

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
13 December 2001 (13.12.2001)

PCT

(10) International Publication Number
WO 01/94542 A2

- (51) International Patent Classification⁷: C12N (72) Inventor: FELDMANN, Richard, J.; 17800 Mill Creek Drive, Derwood, MD 20855-1019 (US).
- (21) International Application Number: PCT/US01/16471 (74) Agent: ZEGER, Jim; 801 North Pitt Street #108, Alexandria, VA 22314 (US).
- (22) International Filing Date: 31 May 2001 (31.05.2001) (81) Designated States (national): AU, CA, IL, JP, MX.
- (25) Filing Language: English (84) Designated States (regional): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).
- (26) Publication Language: English
- (30) Priority Data:
60/208,650 2 June 2000 (02.06.2000) US
09/866,925 30 May 2001 (30.05.2001) US
Published:
— without international search report and to be republished upon receipt of that report
- (71) Applicant: GLOBAL DETERMINANTS, INC. For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.
[US/US]; 17800 Mill Creek Drive, Derwood, MD 20855-1019 (US).



WO 01/94542 A2

(54) Title: ALGORITHMIC DETERMINATION OF FLANKING DNA SEQUENCES THAT CONTROL THE EXPRESSION OF SETS OF GENES IN PROKARYOTIC, ARCHEA AND EUKARYOTIC GENOMES

(57) Abstract: An algorithm has been developed to identify four DNA sequences of 20 bases or more that form a structure called a connectron. Two sequences C1 and C2 are adjacent to each other. These sequences are expressed as RNA in the 3' UTR of some genes in many prokaryotic, archaea and eukaryotic genomes. The other half of a connectron is two DNA sequences T1 and T2 that are on the same chromosome and range in distance from each other by about 1kb to 105kb. The C1 sequence is identical to the T1 sequence and the C2 sequence is identical to the T2 sequence. C1/C2 and T1-T2 can be on different chromosomes. The C1/C2 RNA sequence of the gene transcript finds the two double-stranded DNA sequences T1 and T2. The single-stranded RNA and double-stranded DNA then form a triple-stranded Hoogsteen helix of the RNA/DNA/DNA variety. Because the C1 sequence is adjacent to the C2 sequence, the T1 sequence is made spatially adjacent to the T2 sequence in a compact X-shaped structure. Chromatin particles form as compact 30nm assemblies in the DNA between T1 and T2 thus eliminating the intervening genes from promotion and expression. Connectrons remove sets of genes from expression and thus modulate the behavior of many types of cells.

**ALGORITHMIC DETERMINATION OF FLANKING DNA SEQUENCES
THAT CONTROL THE EXPRESSION OF SETS OF GENES IN
PROKARYOTIC, ARCHEA AND EUKARYOTIC GENOMES**

5 Reference to Related Application

The present application is the subject of Provisional Application Serial No. 60/208,650 filed June 2, 2000 entitled ALGORITHMIC DETERMINATION OF CONNECTRONS FOR THE HIGH LEVEL REGULATION OF GENE EXPRESSION.

10 Introduction

RNA introduced into a cell by a virus is now known to trigger a cellular defense mechanism known as post-transcriptional gene silencing (PTGS). If the viral RNA sequence matches a sequence within the cell's genome the associated genes are turned off or silenced. This phenomenon is also called 'RNA interference' or RNAi. A
15 single-stranded RNA can interact with another single-stranded RNA (known as antisense RNA). The single-stranded RNA can also form a triple-stranded complex with double-stranded DNA. This triple-stranded complex is known as a Hoogsteen helix. This patent application shows how two specific adjacent RNA single-stranded sequences (called C1 and C2 – for Control Sequence 1 and Control Sequence 2)
20 interact with two distant double-stranded DNA sequences (called T1 and T2 – for Target Sequence 1 and Target Sequence 2) to form a tetradic relationship which is called a "connectron". The two distant DNA double-stranded sequences (T1 and T2) must be on the same chromosome in a genome and they must be between about 1kb and 105kb of each other. The adjacent single-stranded RNA sequences (C1/C2) can
25 be on the same or different chromosome as the T1 and T2 sequences. The C1 sequence is identical to the T1 sequence and the C2 sequence is identical to the T2

sequence. The connectron acts to stabilize the double-stranded DNA by allowing 30nm chromatin particles to form. Genes that lie between the T1 and T2 sequences when wrapped up in 30nm chromatin particles are not open to promotion and expression. The connectron (i.e. the tetradic relationship between the T1-T2 sequences and C1/C2 sequences) provides a general explanation for PTGS. A connectron can implemented by RNA sequences, PNA (Peptide Nucleic Acid) sequences or by a zinc-finger DNA Binding Protein (DBP) specific to the T1 and T2 sequences.

Characteristically the adjacent C1/C2 sequences lie in the 3'UTR of a gene. The T1 and T2 sequences do not lie within the translated region of any gene. These sequences "surround" one or more genes. There are, however, T1 and T2 sequence pairs that surround one or more C1/C2 sequences that are not 3'UTR to any gene. These are called "geneless connectrons". There may be promoter sequences that cause the transcription of these 3'UTR sequences.

A computer-based algorithm that is similar to the algorithm used in the US Patent 6,205,404 has been developed to determine the connectron structure of any genome. This algorithm determines the existence of all the connectrons in the genomic DNA. Connectrons exist in prokaryotes, archea, single-celled eukaryotes, multi-celled eukaryotes, plants and higher animals. Connectron relationships exist between prokaryotes and their plasmids. The geneless connectrons provide a possible mechanism for forming a hierarchy of gene expression control that will produce an understanding of cell differentiation and tissue development.

Each connectron is a unique tetrad of sequences. Each connectron changes the expression of the genes between the T1 and T2 sequences. The C1 sequence (which is equivalent to the T1 sequence) and the C2 sequence (which is equivalent to the T2 sequence) are determined by the invention described in this patent application. In general, the tetrad of connectron sequences can be patented because the structure of matter is known and the function of specific gene expression modulation is also known. Gene expression modification can be produced by introducing antisense

RNA or PNA to interact C1/C2 RNA sequences or zinc-finger DBPs to interact with the T1 and T2 sequences. Using connectrons it will be possible to modify cellular and tissue behavior in a very general manner.

5 Examples will be given from different genomes to illustrate that the connectron is a perfectly general and universal concept.

10

Definitions

Double stranded DNA – Watson and Crick showed in 1953 that DNA naturally forms a double-stranded helix. A typical double stranded sequence is

15

5'-TAGAGGAGTACCAC-3'

3'-ATCTCCTCATGGTG-5'

20

Hydrogen Bond - The force between a hydrogen atom and another heavier atom such as Oxygen (O), Nitrogen (N), Phosphorus (P), or Sulfur (S).

Positive strand – The positive strand is normally represented 5' to 3' running left to right as in

25

5'-TAGAGGAGTACCAC-3'

Negative strand – The negative strand is normally represented 5' to 3' running right to left as in

30

3'-ATCTCCTCATGGTG-5'

Single stranded RNA – Either the positive or the negative strand of the double-stranded DNA can be transcribed by the polymerase. In RNA U replaces T.

5 RNA of positive strand sequence 5'-UAGAGGAGUACCAC-3'
RNA of negative strand sequence 5'-GUGGUACUCCUCUA-3'

Antisense RNA – The antisense strand of any RNA sequence is the complement sequence

10

RNA sequence 5'-UAGAGGAGUACCAC-3'
Antisense RNA sequence 3'-AUCUCCUCAUGGUG-5'

Triple Strand Helix – The RNA sequence of a RNA/DNA triple-strand complex is the same as the positive strand of the DNA

15

DNA positive strand 5'-TAGAGGAGTACCAC-3'
DNA negative strand 3'-ATCTCCTCATGGTG-5'
RNA strand 5'-UAGAGGAGUACCAC-3'

20

Promoter – Any region of DNA, that binds proteins which engage the polymerase transcription mechanism.

TATA Box – A region near the 3' end of a promoter with the sequence TATA.

25

mRNA – The RNA produced from the DNA by the polymerase as a result of transcription

Start of transcription – The 3' end of a promoter where the polymerase mechanism begins to transcribe DNA into mRNA.

30

Exon – Any region of mRNA which is used to code for proteins

Intron – Any region of mRNA lying between two exons which is not used to code for proteins. The introns are edited out of the initial RNA transcript to form the mature mRNA.

5

3' UTR – The untranslated 3' end of an mRNA is beyond the end of the last exon. A stop codon in the mRNA causes the ribosome to stop the translation of mRNA into protein.

10

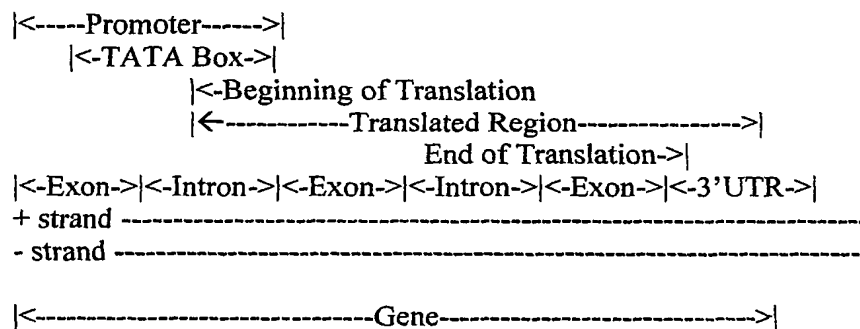
End of translation – The 3' end of the 3'-most exon.

Translated region – Any collection of exons and introns.

15

Gene – Any DNA region that codes for a protein. Introns do not occur in prokaryotic genes and they sometime fail to occur in eukaryotic genes. A typical model of a gene is

20



25

30

Positive strand gene – Any gene in which the features run 5' to 3' on the positive strand

Negative strand gene – Any gene in which the features run 5' to 3' on the negative strand

35

C1 sequence – Any positive or negative strand DNA sequence of 20 bases or more.

The C2 sequence must occur in the same chromosome as the C1 sequence.

C2 sequence – Any positive or negative strand DNA sequence of 20 bases or more.

The C1 sequence must occur in the same chromosome as the C2 sequence.

5

C1/C2 – Any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence

10

T1 sequence – Any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence. The T1 and T2 sequences must be between about 1kb and 105kb apart.

15

T2 sequence – Any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence. The T2 and T1 sequences must be between about 1kb and 105kb apart.

20

Last exon gap or Gap-Distance – The number of bases between the end of transcription and the beginning of the C1/C2 sequence. In prokaryotes and single-celled eukaryotes this gap can range from no bases to 500 bases. In multi-celled eukaryotes the gap can be as large as 10,000 bases.

Poly-adenylation signal – A number of Adenosine (A) bases are added to the mRNA at the end of the 3'UTR.

25

Possible Connectron – Any set of T1, T2 and C1/C2 sequences such that the C1 sequence is identical to the T1 sequence and the C2 sequence is identical to the T2 sequence. The promoter of some gene causes the mRNA of the gene to be expressed. The mRNA is edited to eliminate the introns. The whole mRNA including the 3'UTR can move about in the cell or the nucleus of the cell. The C1/C2 RNA that is part of the 3'UTR moves to the T1 and T2 DNA sequences. A triple-stranded complex of the DNA and the RNA forms such that the C1 sequence forms hydrogen bonds with the T1 sequence and the C2 sequence forms hydrogen bonds with the T2 sequence.

30

Because the C1 sequence is adjacent to the C2 sequence, the T1 sequence is brought physically close to the T2 sequence. This produces a loop of between about 1kb and 105kb in the DNA. Histone proteins reduce the length of the DNA by binding 200 bases. Histone/DNA complexes form six-fold symmetry chromatin assemblies. The diameter of the chromatin assemblies is approximately 30nm.

Real Connectron – Any Possible Connectron which is within the Gap-Distance of some gene

Homologous connectron – The T1 sequence and the T2 sequence are on the same chromosome as the C1/C2 sequence

Heterologous connectron – The T1 sequence and the T2 sequence are on a chromosome different from chromosome of the C1/C2 sequence

Permanent connectron – Any C1/C2 sequence, which is 3' UTR to some gene that is not surrounded by any T1 and T2 sequence pairs

Transient connectron – Any C1/C2 sequence, which is 3' UTR to some gene that is surrounded by one or more T1 and T2 sequence pairs

Self-limiting connectron – Any C1/C2 sequence which is 3'UTR to some gene that is surrounded by the T1 and T2 sequences such that C1=T1 and C2=T2

Geneless connectron - Any C1/C2 sequence which is not 3'UTR to some gene but is surrounded by some T1 and T2. A promoter may lie 5' to the C1/C2 sequence.

Bidirectionality of Connectron Excitation – A C1/C2 short loop on one strand selects a T1-T2 long loop pair on the same or the opposite strand. The C1/C2 short loop has a complementary C1'/C2' sequence on the opposite strand. Similarly the T1-T2 long loop pair has a complementary long loop pair T1'-T2'. Wherever a C1/C2, T1-T2 tetrad exists there is a complementary C1'/C2', T1'-T2' tetrad. The C1/C2 short loop

can be transcribed as a 3'UTR to a gene on the same strand. The C1'/C2' short loop which is on the strand opposite to the C1/C2 short loop can also can be transcribed as a 3'UTR to a gene on the same strand. There are four possible models of action

T1 T2 gene - C1/C2

+ strand -----

- strand -----

5

T1 T2

+ strand -----

- strand -----

C2/C1 - gene

10

+ strand -----

- strand -----

T2' T1' C2'/C1' - gene

15

gene - C1'/C2'

+ strand -----

- strand -----

T2' T1'

20

Of course, the short loops and the long loops do not have to be on the same chromosome.

Hierarchy of connectron action – When a C1/C2 is expressed it forms a T1-T2 loop by forming a connectron. The C1/C2 sequence does not have to be on the same chromosome as the T1 and T2 sequences. This provides a way of causing interaction between chromosomes. When the T1-T2 loop forms, any genes in that loop region which had been expressing C1/C2 sequences in their 3'UTRs, now cease expressing the C1/C2 sequences. The connectrons formed by these C1/C2 sequences will cease to exist after some time thus opening up the genes inside the respective T1-T2 loops

25

30

to expression. The hierarchy of connectron action is alternates between repression and expression. The connectron hierarchies can be of any depth.

5 One-to-Many connectron action – One C1/C2 sequence can form connectrons in many different places on many different chromosomes. The only requirement is that C1=T1 and C2=T2. This makes it possible for one expression event to control the expression of many genes on different chromosomes.

10 Many-to-One connectron action – C1/C2s that come from many different places on many different chromosomes can form a connectron for a specific T1-T2 sequence pair. The only requirement is that C1=T1 and C2=T2. This makes it possible for many different expression events to control the expression of one set of genes on a particular chromosome.

15 Many-to-Many connectron action – The arrangement of C1/C2s and T1-T2s across chromosomes can form a complex web of gene expression control relationships.

20 Percentage of the Genome Regulated by Connectrons – Since the connectrons for a sequenced genome can be calculated, the percentage of the genome that is open to connectron regulation can be known.

25 Emergent Property – The network of connectrons in any genome emerges from a knowledge of the complete DNA sequence of the genome. Because both the C1/C2 sequences and the T1-T2 sequences can be any place in the genome, the whole genomic sequence must be known before all the connectrons can be determined.

30 Paradigm Shift – For the past fifty years since the discovery by Watson and Crick of the double-helical nature of DNA, the reigning paradigm for scientific discovery has been the study of one gene and its effects on the behavior of a cell. The advent of genomic sequencing and this invention of connectrons that emerge from the whole genome will produce a shift in the way scientists view biological systems and the way they formulate and execute experiments. The many-to-many relationships between

the connectrons means that there are many ways in which the expression of a set of genes can be modulated. The multiplicity of control pathways means produces a system stability that makes it possible for biological systems to be stable for long periods of evolutionary time. The thinking that goes into formulating scientific experiments will have to change to accommodate the changes in understanding that will be induced by the application and extension of this patent application.

Hierarchy of DNA Structuring – The DNA of a cell's genome is structured in a hierarchy of six levels. Figures 1, 2 and 3 have been adapted from The Molecular Biology of the Cell by Alberts, Bray, Lewis, Raff, Roberts and Watson [third edition pages 354, 345 and 348]. As shown in figure 1, the double stranded DNA is level 1. The double-stranded DNA is wrapped around histone proteins to form a chromatin particle that is level 2 of the hierarchy. Level 2 is described as "beads-on-a-string" in figure 1. The chromatin particles are packed in a six-fold symmetry as shown in figure 2a and figure 2b. These six-fold assemblies have a diameter of 30 nm. Each 30 nm assembly contains from 18 (i.e. $6 * 3$) to 30 (i.e. $6 * 5$) chromatin particles. The 30 nm assemblies aggregate into large loops which range in length from 5,000 bases to 100,000 bases of DNA. The size of these large loops as shown in figure 1 is approximately 300 nm. These large loops constitute level 4 of the structuring hierarchy. As shown in figure 1, level 5 of the DNA structuring hierarchy many large loops are condensed to form a structure which is approximately 700 nm in diameter. The complete chromosome that constitutes level 6 of the hierarchy is composed of two very long sections of level 5 DNA.

Model of Chromatin Structure – The level 4 structure of DNA as shown in figure 1 ranges in length from 5,000 to 105,000 bases of DNA. Figure 3 shows that proteins are thought to connect portions of the long loops formed by the 30 nm particles to form a chromosome axis. These condensed long loops are described as chromomeres in The Molecular Biology of the Cell.

Prior Art

5 The chromomere model of DNA structuring was presented by N. A Resnik, et al.[1] and is based on electron microscopic data. There are more recent papers studying a variety of genomes with electron microscopy but no equivalent study of chromomeres has been done on a fully sequenced genome.

10 A recent News Feature in Nature by T. Gura [2] described the discovery of post-transcriptional gene silencing in which viral RNA interacts with the transcribed RNA of the cell to silence the expression of genes. This article describes experiments in *C. elegans* and *D. megalomaster* in which RNA that is complementary to mRNA introduced into a cell. This "antisense" RNA has the effect of turning off the expression of one or more genes. The introduced complementary RNA produces an "RNA interference" called RNAi.

15 Thomas Werner and his colleagues at Genomatix in Munich, Germany have developed an approach to understanding what they call "Matrix Attachment Region" (MAR). Figure 5 shows their interpretation of the structure of DNA surrounding a gene. The following description of the MAR is copied from the Genomatix web site

20 "Matrix Attachment Regions (MARs) MARs are sequence regions that are responsible for the attachment of genomic DNA to the nuclear matrix or scaffold. Transcription absolutely requires anchorage of genomic DNA to the nuclear matrix.

Functional features of MARs:

25 Anchoring of regulatory elements like promoters and enhancers to the nuclear matrix.

Ensuring long term activity of promoters and enhancers in chromatin.

Insulation, rendering a functional domain insensitive to position effects.

Genomatix is conducting a research project to define and detect MARs by computer-analysis.”

Brief Description of the Objects of the Invention

5 An object of the invention is to provide a method of identifying DNA sequences that control the expression of different collections of genes in a genome comprising, detecting selected DNA sequences adjacent to some genes excluding exons and introns.

10 An object of the invention is to provide a method of identifying DNA sequences that control the expression of different collections of genes comprising, detecting, by computer, one or more pairs of non-adjacent DNA sequences to which are bound to two RNA sequences.

15 An object of the invention is to provide a method of identifying DNA sequences that control the expression of different collections of genes in a genome comprising detecting changes in connectron behavior in the genome.

20 An object of the invention is to provide a method of modifying the expression of different gene collections in a genome, comprising detecting changes in connectron behavior as a result of an exogenous stimulus.

25 An object of the invention is to provide a method of detecting where and when new genes are being integrated into a host genome comprising detecting the connectrons in said host genome.

 An object of the invention is to provide a method of detecting the expression effect of different gene collections in a given body comprising detecting the back and forth flow of connectrons between the chromosomes thereof.

An object of the invention is to provide a method of modifying a given body comprising modifying the connectron organization therein.

5 An object of the invention is to provide a method of detecting connectron control and target sequences in a given genome comprising:

determining the base composition of said genome,
determining one or more sites of control sequence organization, and/or
determining one or more sites of target application.

10 An object of the invention is to provide a method of determining the response of a cell in any tissue to changes in the cell's environment and/or genetic composition comprising providing a complete genomic DNA sequence for the organism and determining the effect of changes in connectrons due to application of a given
15 exogenous stimulus to the genome.

An object of the invention is to provide a method of determining in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes, the tetradic relationship $T1=C1$ and $T2=C2$ where T1 and T2 are DNA sequences 20 or more bases in length,
20 where the C1 sequence is adjacent to the C2 sequence, where the T1 and T2 sequences are on the same chromosome, and where the C1/C2 sequences are on the same chromosome as T1 and T2 or where the C1/C2 sequences are on a chromosome different from T1 and T2, wherein:

25 C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

30 C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

5 T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

10 T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

15 An object of the invention is to provide a method of determining in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes, the connectron relationship that permits many different C1/C2 short loops to control the existence of a T1-T2 long loop and wherein said C1/C2 short lops can be on the same chromosome or on different chromosomes from the T1-T2 long loop, wherein:

20 C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

25 C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

30 T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

5 An object of the invention is to provide a method of determining in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes, the connectron relationship that permits one C1/C2 short loop to control the existence of many T1-T2 long loops, the C1/C2 short loop can be on the same chromosome or on different chromosomes from the T1-T2 long loops, wherein:

10

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

15

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

20

C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

25

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

30

An object of the invention is to provide a method of determining in the connectron relationships between prokaryotes and their plasmids wherein said connectrons implement a control mechanism between the two genomes that makes it possible from

them to form a symbiotic relationship, and in the case of *D. radiodurans* the relationship is not symmetric, and the *D. radiodurans* genome sends C1/C2 short loops to the MP1 plasmid, wherein:

5 C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

10 C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

15 C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

20 T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

25 An object of the invention is to provide a method of determining that connectron relationships that exist in plant and higher animals.

30 An object of the invention is to provide a method of determining in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes, the connectron relationship that permits one C1/C2 short loop to control the existence of one or more T1-T2 long loops without being subject to any expression controls other than those of the gene to which the C1/C2 is 3'UTR, wherein:

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

5 C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

10 C1/C2 - any positive or negative strand DNA sequence of 540 or more bases such that the C1 sequence is adjacent to the C2 sequence,

15 T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart,

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart, and

20 3'UTR - untranslated 3' end of an mRNA is beyond the end of the last exon, a stop codon in the mRNA causes the ribosome to stop the translation of mRNA into protein.

25 An object of the invention is to provide a method of determining in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes, the connectron relationship that permits one C1/C2 short loop to control the existence of one or more T1-T2 long loops such that this C1/C2 short loop is itself subject to expression control by another T1-T2 long loop which surrounds it, wherein:

30 C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

5

C1/C2 - any positive or negative strand DNA sequence of 540 or more bases such that the C1 sequence is adjacent to the C2 sequence,

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

10

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

15

An object of the invention is to provide a method of determining in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes, the connectron relationship that permits one C1/C2 short loop to control the existence of the T1-T2 long loop that surrounds it, wherein:

20

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

25

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

30

C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

5 T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

10 An object of the invention is to provide a method of determining the connectron relationships that do not have any genes within the T1-T2 long loop, wherein:

T1 sequence is any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, and

15 T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, and the T2 or T1 sequences must be between about 1kb and 105kb apart.

20 An object of the invention is to provide a method of determining the geneless connectron relationship where one C1/C2 short loop controls the existence of many geneless T1-T2 long loops, wherein:

25 C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

30 C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

5

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

Description of the Drawings and Tables

The above and other objects, advantages and features of the invention will become more apparent when considered with the following specification and accompanying drawings and tables wherein:

Figure 1 DNA is structured in six levels of increasing condensation. Double stranded DNA is level 1. Two turns of DNA are wrapped about each chromatin particle at level 2. The chromatin particles which each containing 200 base pairs form into 30 nm particles at level 3. The 30 nm particles form into large loops with an approximate dimension of 300 nm at level 4. Metaphase chromosomes form a condensed structure with an approximate dimension of 700 nm at level 5. An entire metaphase chromosome has a width of approximately 1400 nm at level 6. The large loops at level 4 of the DNA structuring are thought to have between 20,000 (20 kb) and 100,000 (100 kb) base pairs.

The Molecular Biology of the Cell by Alberts, Bray, Lewis, Raff, Roberts and Watson, 3rd. ed. , Garland Publishing, Inc., New York, 1994, p. 354

Figure 2 (a) Chromatin DNA forms into a six-fold symmetry 30nm particles.
(b) The six-fold symmetry 30nm particles form a linear chain with a varying number of repeat units.

The Molecular Biology of the Cell by Alberts, Bray, Lewis, Raff, Roberts and Watson , 3rd. ed. , Garland Publishing, Inc., New York, 1994, p. 345

Figure 3 Long loops of 30nm particles are thought to be closed at the bottom of the loop by proteins.

5 The Molecular Biology of the Cell by Alberts, Bray, Lewis, Raff, Roberts and Watson, 3rd. ed. , Garland Publishing, Inc., New York, 1994, p. 348

Figure 4 (a) Transcription and Editing. (b) Movement of the RNA through the Nucleus. (c) Connectron Formation

10 Figure 5 Overview of schematic organization of a typical transcriptionally active chromosomal loop.

From http://genomatix.gsf.de/func_genomics/functional_genomics.html

15 Table 1 Connectron Properties for Prokaryotic, Archea and Eukaryotic Genomes

Table 2 Yeast Inter-Chromosomal Connectron Distribution

- Figure 6 Genome size plotted as a log-log function of the Number of Connectrons
- Figure 7 Number of Sequence Instances plotted as a function of the Number of Fragments
- 5 Figure 8 Level 0 – The overall view of the algorithm
- Figure 9 Level 1 – Process Flow of the Algorithm
- Figure 10 Level 2a – two pages – Process Genome into Blocking Fragment File
- Figure 11 Level 2b – two pages – Compute the Connectrons for a Genome
- Figure 12 Level 2c – two pages – Analyze Possible Connectrons
- 10 Figure 13 Level 3a – Setup Genome Usage Memory
- Figure 14 Level 3b – Find DBP-Size Blocking File for T1-Window
- Figure 15 Level 1 – Find DBP-Size Blocking File for T2-Window
- Figure 16 Level 2a – two pages – Find C1/C2 Entries
- Figure 17 Level 2b – two pages – Scan Genome Usage Memory for Potential
15 Connectrons

Description of the Invention

A connectron is a relationship among four DNA sequences. Each sequence must be at least 20 bases long. There is a report by Sharp and Zamore [3] that RNA sequences of "about length 25" are important as sources of RNAi. 27 bases were actually used as the minimum length of each of the sequences. The T1 sequence is on one strand of some chromosome in a genome. The T2 sequence is on the same strand of the same chromosome as the T1 sequence. The T1 and T2 sequences (which are each at least 20 bases in length) must be at least 5,000 bases distant from each other but they can not be more than 105,000 bases distant from each other. The C1 sequence and the C2 sequence (which are each at least 20 bases in length) are adjacent to each other on some strand of some chromosome in the genome. The C1/C2 sequences – called the "short loop" – can be on the same strand as the T1 and T2 sequences or they can be on the opposite strand. The C1/C2 sequences of the short loop can be on the same chromosome as the T1 and T2 sequences but they can also be on a different chromosome in the genome. When a genome has only one chromosome, then the point is moot. Many genomes, of course, have several chromosomes. The C1 sequence is identical to the T1 sequence and the C2 sequence is identical to the T2 sequence.

The C1/C2 sequence must be on the same strand as a gene, either be directly adjacent to the gene (i.e. a gap of 0 bases) for prokaryotic genomes or at this time be within 10,000 bases for eukaryotic genomes. The size of the gap between the end of the gene and the beginning of the C1/C2 sequence is a variable. The C1/C2 short loop is expressed as the 3'UTR (Un-Translated Region) of the gene. In the case of prokaryotic genes that do not normally have introns, the whole mRNA becomes the active species for connectron formation. In the case of eukaryotic genes, the whole transcript is the active species for connectron formation upon editing of the transcript to eliminate the introns. The whole transcript then can move about in the cytoplasm of prokaryotic cells or the nucleus of eukaryotic cells. Since the C1 sequence is equivalent to the T1 sequence and the C2 sequence is equivalent to the T2 sequence,

the C1 RNA can form a Hoogsteen triple-stranded RNA/DNA/DNA helix with the double-stranded T1 sequence. Similarly the C2 RNA can form a Hoogsteen triple-stranded RNA/DNA/DNA helix with the double-stranded T2 sequence. Because the C1 sequence and the C2 sequence are adjacent to each other, the C1/T2 RNA/DNA/DNA Hoogsteen triple helix is brought into physical adjacency to the C2/T2 RNA/DNA/DNA Hoogsteen triple helix. RNA/DNA/DNA hybrid helices are the most stable form of triple helix. RNA double helices, DNA double helices, RNA triple helices and DNA triple helices are all significantly less stable than a RNA/double-stranded DNA triple helix. The stable physical adjacency of the two triple-stranded Hoogsteen helices ensures that the long loop of double-stranded DNA between the T1 sequence and the T2 sequence can then be structured into 30 nm chromatin particles as shown in level 4 of figure 1. The genes on either strand of the DNA between the T1 sequence and the T2 sequence when they are structured into the 30 nm chromatin particles are not open to promotion and expression.

The tetradic relationship between the T1 and T2 sequences that form the long loop and the C1/C2 sequences that form the short loop are called a connectron. The name "connectron" was suggested by J. David Rawn Ph.D. of Towson University. A connectron is possible if the T1, T2, C1 and C2 sequences exist. A connectron is real if the C1/C2 short loop sequence is adjacent to an expressible gene. If the expression of the adjacent gene is inside one or more T1 - T2 long loops then this connectron is said to be transient. If the adjacent gene is not inside any possible T1-T2 long loop then the connectron is said to be permanent. If a connectron is inside of a T1-T2 long loop that has the same sequences (i.e. T1 is really equal to C1 and T2 is really equal to C2) then the connectron is said to be self-limiting. This is true because once the C1/C2 sequence is expressed it forms the T1-T2 long loop that then shuts off the expression of the gene adjacent to the C1/C2 sequence. Self-limiting connectrons can also be called "spike" connectrons since they generate a short-duration spike of the C1/C2 short loop sequence. If a T1-T2 long loop does not contain any genes but it contains C1/C2 short loop sequences then this type of connectrons is said to be geneless. The C1/C2 short loops within a geneless T1-T2 long loop can, of course, control the expression of genes.

The physical existence and lifetimes of the connectrons must be proved by molecular biological experimentation. This physical experimental process, however, is logically quite separate from the computational experimentation that have been conducted from June of 1999 to May of 2001. The computational search for the existence of connectrons has been extremely positive. These computations have shown that connectrons exist in prokaryotes, in archea, between prokaryotes and their plasmids, in single-celled eukaryotes, in multi-celled eukaryotes, in plants, in higher animals and in humans. All of these features and properties are described in the claims section that follows.

The connectron invention is very powerful. It depends only on sequence equivalency. The minimum length of the four sequences seems to be about 20 bases. In the calculations shown in this patent application, 27 bases have been used as a minimum. The Nature News Feature [1] says that other scientists have found RNA sequences of length about 25 that have interesting gene silencing properties. The Nature article does not give any mechanism. Because of my algorithm and its use on a variety of genomes, this patent application provides the computational proof that a particular mechanism is highly probable. The connectron invention provides an explanation for how communication occurs with a chromosome as well as between chromosomes in genomes that have more than one chromosome. Since each T1-T2 long loop can contain one or more genes, the connectron invention provides a mechanism for turning on and turning off sets of genes simultaneously. In time, the connectron invention will provide an explanation for how differentiation of how one cell's behavior differs from the behavior of another adjacent cell. It is already clear from the computational experiments that have been made on *S. cervesiae*, *C. elegans* and *D. megalomaster* that the number of geneless connectrons increases dramatically as evolution proceeds from single-celled eukaryotes (i.e. *S. cervesiae*) to 1,000 cell eukaryotes (i.e. *C. elegans*) to visible creatures (i.e. *D. megalomaster*). The extension of this evolutionary progress to plants (i.e. *A. thaliana*) for which only three chromosomes are sequenced and humans (i.e. *H. sapiens*) for which only one chromosome is completely sequenced. Although the complete human genome was

published in Nature and Science in February of 2001, the NIH-sponsored genomic sequencing results are available for about 1/3 of the bases in the whole genome. The human genomic sequence determined by Celera Genomics, Inc. is available only by subscription. Table 1 shows how the genome size, the number of genes, the number of
5 gene-containing and geneless connectrons and the percentage of genes controlled are related in many different genomes.

The C1/C2 short loops originate on one chromosome. The T1-T2 long loops can be on the same or different chromosomes. Table 2 which is for yeast (*S. cerevisiae*) is a
10 square matrix of how many C1/C2 short loops on a given chromosome are sent to form T1-T2 long loops on other chromosomes. The diagonal of this matrix shows that many chromosomes send connectrons to themselves. The striking feature of this particular table is that chromosome 6 only sends connectrons to chromosome 12 but that it receives connectrons from chromosomes 4,5,7,10,12,13,15 and 16.

Any tetrad of connectron sequences (i.e. the T1, T2, C1 and C2 sequences) as well as the fact of the adjacency of the C1/C2 short loop sequence to the transcribing gene can be patented because the content of matter and the utility can be exactly described. The utility of a connectron is that the T1-T2 long loop shuts off the expression of the
15 genes that lie between the T1 sequence and the T2 sequence. In the case of geneless connectrons, the utility is of a higher level in that the C1/C2 short loops contained in the higher-level geneless T1-T2 long loop, eventually form other lower-level T1-T2 long loops around a set of genes.

The invention of connectrons comes at a particularly important time in biological discovery. The geneless connectrons make a many-to-many hierarchical control mechanism possible. It is already clear from the determination of the connectrons for *C. elegans* and *D. megalomaster* that there are as many or more geneless connectrons than there are genes. It has been clear for some time that the number of genes in a
20 genome is not particularly correlated with the size of the genome. Figure 6 shows that the size of a genome is roughly linearly correlated with the number of connectrons.

The connectron invention can be used to generate a model of behavior in any cell. The simulation of connectron behavior in different genomes will be the subject of another patent application.

5

The connectron invention provides for a rational exploitation of the information contained in the raw genomic DNA sequence by forming a hierarchy of relationships between geneless connectrons, transient connectrons, permanent connectrons, self-limiting connectrons and the expression of genes.

10

Detailed Description of the Invention

The algorithm for the determination of connectrons in any genome or any genome fragment is represented in the following flow diagrams. The Level 0 diagram in figure 8 shows the general relationships in a digital computer. The central processor of the digital computer uses the computer program to take genome descriptors, the genomic DNA sequences and the tables of gene features to produce a file of blocking fragments and a file of the optimal connectrons for the genome. The printer serves to make hard copies of the files and this patent application. The level 1 diagram in figure 9 shows the three essential steps in the determination of connectrons. The genome is first processed into a blocking fragment file. Then the blocking fragments are used to compute the connectrons for the genome. Finally the potential connectrons are analyzed to determine if the C1/C2 sequences are in the 3'UTR of a gene. The level 2a diagram in figure 10 shows the steps required for the processing of the genome into a file of blocking fragments. The genomic DNA sequence is decomposed into 27-base frames for both the positive and negative strands. These fragments are written to the unsorted fragment file. The fragment file is then sorted is then read and formed into groups of equivalent sequences. The (.blk) file contains the sequence and a pointer to the (.gpnr) file which contains the pointers to the position of the fragments in the genomes. The position in the genome includes the chromosome number, the position in the chromosome and the strand (i.e. positive and negative). A sample of these files follows

Sample of the (.blk) file for *S. cerevisiae*

27-base fragment	Number of instances	Pointer to (.gpnr) file
11111111111111111111111111111111	0	1
111111123244233313332443414	1	2
111111141113443133314333341	2	4
111111232442333133324434141	1	5
111111323311133323144423444	2	7
111111332213331341414443413	2	9
111111333444112343412323243	1	10

	111111333444113343412323243	9	19
	111111411134431333143333414	2	21
	111111443223134142124434124	2	23
	111112223234344444443144442	2	25
5	111112244123441122214421213	8	33
	111112311241114344334134431	2	35
	111112324423331333244341414	1	36
	111112344232231344242234342	1	37
	111112433444244421144134211	1	38
10	111112444311313442332142224	1	39
	111113131241131114424413231	1	40
	111113143332344311113133411	1	41
	111113233111333231444234441	2	43

15 In fragments above 1=G, 2=C, 3=A, 4=T

Sample of the (.gptr) file for *S. cerevisiae*

There are 16 chromosomes in *S. cerevisiae*

20	Item	Chromosome	Position in Chromosome	Direction
	1	0	0	0
25	2	4	11137	1
	3	12	467619	1
	4	12	458482	1
	5	4	11138	1
	6	12	465759	2
30	7	12	456622	1
	8	1	219366	1
	9	8	539978	1
	10	14	522451	1
	11	4	1099073	1
35	12	4	1210003	1
	13	7	539068	1
	14	12	654136	1
	15	12	596455	1
	16	15	121016	1
40	17	15	598127	2
	18	16	847724	1
	19	16	59765	1
	20	12	467620	1
	21	12	458483	1
45	22	12	461657	1
	23	12	452520	1
	24	13	838006	1

	25	15	288270	1
	26	4	83593	1
	27	4	992867	1
	28	6	162265	1
5	29	7	845687	1
	30	10	531560	2
	31	15	282208	1
	32	16	860418	1
	33	16	572308	1
10	34	12	465992	1
	35	12	456855	1
	36	4	11139	1
	37	8	89343	1
	38	4	10302	1
15	39	1	19894	2
	40	16	9311	1
	41	10	735203	1
	42	12	465760	1
20	43	12	456623	1

In direction column above 1=positive strand, 2=negative strand

The level 2b diagram in figure 11 shows the computation of the connectrons. The genome descriptors consist of the number and length of the chromosomes. The algorithm uses an array that represents several facts about each base position in the genome. The level 3a diagram in figure 13 shows the setup of the Genome-Usage memory. The gene features are used to prevent the region of the genome that codes for proteins from being used for the connectron sequences (i.e. the T1s, the T2s, the C1s and the C2s). In the level 2a diagram of figure 10, the algorithm steps through each chromosome and within each chromosome through each base position looking for acceptable T1-windows of 27 bases. A T1-window can be used to form a connectron relationship if there are two or more instances of this fragment in the blocking fragment file. The computation in the level 3b diagram of figure 14 determines if the T1-window is acceptable or not. Once an acceptable T1-window is found, the algorithm (in the level 2a diagram of figure 10) looks for acceptable T2-window positions that lie between 5,000 and 105,000 bases from the T1-window. The computation for determining acceptable T2-window positions is done in the level 3c diagram of figure 15. Once a pair of T1 and T2 window positions are found, the

algorithm looks among the instances of these T1 and T2 sequences for a pair of sequences C1 and C2 that lie within 200 bases of each other on the same chromosome. The computation for determining acceptable C1/C2 windows is shown in the level 3d diagram in figure 16. In the level 3e diagram of figure 17 the Genome-Usage memory is scanned for the Possible-Connectrons. In the level 2c diagram of figure 12 the Possible-Connectrons are scanned to determine if the C1/C2 sequences are within the Gap-Distance of a gene on either the positive or the negative strands. The Real-Connectrons are then written out in several different files including the descriptions in the claims section.

10

Examples

The algorithm for the determination of optimal connectrons has been applied to a number of different publicly available genomes. The connectron is a tetradic relationship between four sequence elements – T1, T2, C1 and C2. The claims presented in this section are written by the program NearGene that implements the flow diagram Level 2c of figure 12. The examples are written a uniform type of English. Each example contains some or all of the following elements

10 Name of genome
 Description of T1
 Length of T1-T2 loop
 The chromosome on which the T1-T2 loop exists
 The identifier number within the genome of the T1 sequence
 15 The T1 sequence
 Description of T2
 The identifier number within the genome of the T2 sequence
 The T2 sequence
 A list of genes whose expression is controlled by the T1-T2 loop
 20 The common names of the genes as obtained from the NCBI gene feature file
 (.ptt)
 A list of C1/C2 short loops whose expression if controlled by the T1-T2 loop
 The chromosome on which the C1/C2 short loop exists
 The common name of the gene which expresses the C1/C2 short loop as an
 25 RNA
 The sequence of the C1/C2 short loop
 A list of C1/C2 short loops that control the formation of the T1-T2 loop
 The chromosome on which the C1/C2 short loop exists
 The common name of the gene which expresses the C1/C2 short loop as an
 30 RNA
 The sequence of the C1/C2 short loop

The match between the C1/C2 sequence and the T1 sequence

The match between the C1/C2 sequence and the T2 sequence

5

The uniform descriptions make it possible to rapidly comprehend the specifics in each example.

When a sequence element is very long a series of four dots has been inserted between the beginning and ending sequence groups. A variable number of bases have been deleted.

10

Index of Pages for Connectron Samples

Page 39

- 5 1. Connectrons occur in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes.

Page 57

- 10 2. Many Connectrons control the expression of one set of genes in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes.

Page 83

- 15 3. One connectron controls the expression of many sets of genes in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes.

Page 107

4. Connectrons occur between prokaryotes and their plasmids.

Page 117

- 20 5. Connectrons occur in plants and higher animals

Page 126

6. Permanent connectrons exist in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes.

Page 135

- 25 7. Transient connectrons exist in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes.

Page 152

- 30 8. Self-limiting connectrons occur in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes

Page 164

9. Geneless connectrons exist in single-celled and multi-celled eukaryotes

5

Page 174

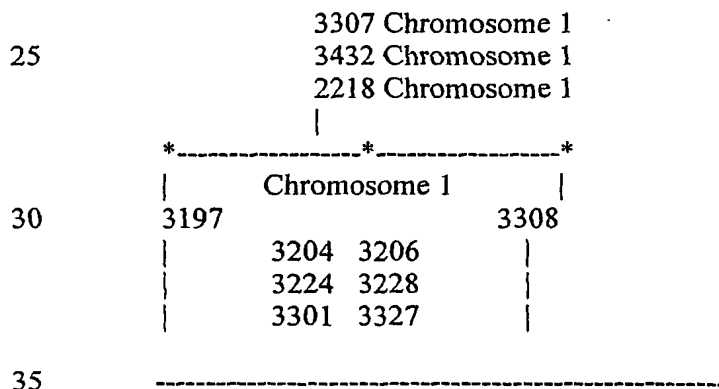
10. One connectron controls many geneless connectrons in single-celled and multi-celled eukaryotes

1. Connectrons occur in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes.

Connectrons exist as tetradic relationships where the sequence T1 is equivalent to the sequence C1 (written T1=C1) and where the sequence T2 equals the sequence C2 (written T2=C2) where T1 and T2 are DNA sequences 20 or more bases in length, where the C1 sequence is adjacent to the C2 sequence, where the T1 and T2 sequences are on the same chromosome, and where the C1/C2 sequences are on the same chromosome as T1 and T2 or where the C1/C2 sequences are on a chromosome different from T1 and T2. The connectron relationship has been found to exist in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes.

Example of a prokaryote connectron – E. coli

In this example the existence of the T1-T2 (3197-3308) long loop is controlled by three C1/C2 short loops (3307, 3432 and 2218). The T1-T2 long loop controls the expression of 64 genes on chromosome 1 in addition to six C1/C2 (3204, 3206, 3223, 3228, 3301 and 3327) short loops. The C1/C2 short loop 3327 lies outside the range of the T1-T2 long loop (3197-3308) but this C1/C2 is expressed as a 3'UTR to the gene hemG that is within the range of the T1-T2 long loop.



Connectron control elements for chromosome 1 of the E. coli genome

A double stranded DNA loop of length 93.542 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 3197. This T1 control element has the DNA sequence

AAAAAATGCGCGGTCAGAAAATTATTTTAAATTCCTCTTGTCAGGCCGG
AATAACTCCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACG
CCGCCGGGTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAA
ATAAATGCTTGACTCTGTAGCGGGAA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 3308. This T2 control element has the DNA sequence

TAAATTCCTCTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTG
ACACGGAACAACGGCAAACACGCCCGGGTCAGCGGGGTTCTCCTGAG
AACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCGGGAAG
GCGTATTATGCACACCCCGCGCCGCT

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

rrsC	gltU	rrlC	rrfC	aspT	trpT	yifA	yifE	yifB
ilvL	ilvG_1	ilvM	ilvE	ilvD	ilvA	ilvY	ilvC	ppiC
b3776	rep	gppA	rhlB	trxA	rhoL	rho	rfe	wzzE
wecB	rffH	wecD	wecE	wzxE	yifM_2	wecG	yifK	
argX	hisR	leuT	proM	aslB	aslA	hemY	hemX	
hemD	cyaA	cyaY	b3808	dapF	uvrD	b3814	corA	
yigF	yigG	rarD	yigI	pIdA	recQ	yigJ	yigK	pIdB
yigL	yigM	metR	metE	ysgA	udp	yigN	ubiE	yigP

b3836 yigU yigW_1 rfaH yigC ubiB fadA fadB
 pepQ trkH hemG

5 This long T1/T2 double stranded DNA loop modulates the expression of the
 following C1/C2 short loops

10 A C1/C2 short loop on chromosome 1 whose identifier is 3204 controls the
 expression of the genes of one or more other T1/T2 long loops. This C1/C2 short
 loop is expressed as a RNA single strand that is 3'UTR to the gene rrsC and has the
 DNA sequence

15 GATGTGCCCAGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGG
 CGACGATCCCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAAGT
 ACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATG
 GCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAA

20 A C1/C2 short loop on chromosome 1 whose identifier is 3206 controls the
 expression of the genes of one or more other T1/T2 long loops. This C1/C2 short
 loop is expressed as a RNA single strand that is 3'UTR to the gene rrsC and has the
 DNA sequence

25 GTCCCCTTCGTCTAGAGGCCAGGACACCGCCCTTTCACGGCGGTAACAG
 GGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTC
 ACCTGCCTTAATATCTCAAACTCATCTTCGGGTGATGTTTGAGATATTTG
 CTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAA

30 A C1/C2 short loop on chromosome 1 whose identifier is 3223 controls the
 expression of the genes of one or more other T1/T2 long loops. This C1/C2 short
 loop is expressed as a RNA single strand that is 3'UTR to the gene rrlC and has the
 DNA sequence

GCTGAAGTAGGTCCCAAGGGTATGGCTGTTTCGCCATTTAAAGTGGTACGC
GAGCTGGGTTTAGAACGTCGTGAGACAGTTTCGGTCCCTATCTGCCGTGGG
CGCTGGAGAACTGAGGGGGGCTGCTCCTAGTACGAGAGGACCGGAGTGG
ACGCATCACTGGTGTTTCGGGTTGTCATGCCAATGGCA

5

A C1/C2 short loop on chromosome 1 whose identifier is 3225 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene rrlC and has the DNA sequence

10

AAACAGAATTTGCCTGGCGGCCGTAGCGCGGTGGTCCCACCTGACCCCAT
GCCGAAGTCAGAAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTC
CCCATGCGAGAGTAGGGAAGTCCAGGCATCAAATTAAGCAGTA

15

A C1/C2 short loop on chromosome 1 whose identifier is 3228 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene rrfC and has the DNA sequence

20

GGTCATAAAACCGGTGGTTGTAAAAGAATTCGGTGGAGCGGTAGTTCAGT
CGGTTAGAATACCTGCCTGTCACGCAGGGGGTTCGCGGGTTCGAGTCCCGT
CCGTTCCGCCAC

25

A C1/C2 short loop on chromosome 1 whose identifier is 3301 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ubiB and has the DNA sequence

30

TTATCGTGCCTACAAATAGTCCGAACCGTAGGCCGGATAAGGCGTTTACG
CCGCATC

A C1/C2 short loop on chromosome 1 whose identifier is 3307 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene fadA and has the DNA sequence

5

TGCCGGATGCGGCGTAAACGCCTTATCCGGCCTACGGTTCGGACTATTTGT
AGGCA

10

A C1/C2 short loop on chromosome 1 whose identifier is 3327 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene hemG and has the DNA sequence

15

AAAAAATGCGCGGTCAGAAAATTATTTTAAATTTCTCTTGTCAGGCCGG
AATAACTCCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACG
CCGCCGGGTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAA
ATAAATGCTTGACTCTGTAGCGGGAAGGCGTATTATG...CCCGTCACACCA
TGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGGCGCT
TACCACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAACCGTA
GGGGAACCTGCGGTTGGATCACCTCCTTACCTTAAAGAAGCGTTCTTTG

20

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

25

A C1/C2 short loop on chromosome 1 whose identifier is 3307 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene hemG and has the DNA sequence

30

AAAAAATGCGCGGTCAGAAAATTATTTTAAATTTCTCTTGTCAGGCCGG
AATAACTCCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACG
CCGCCGGGTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAA
ATAAATGCTTGACTCTGTAGCGGGAAGGCGTATTATG...CCCGTCACACCA

TGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCGCT
TACCACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAACCGTA
GGGGAACCTGCGGTTGGATCACCTCCTTACCTTAAAGAAGCGTTCTTTG

5 The match between the T1 sequence and the C1/C2 sequence is

AAAAAATGCGCGGTCAGAAAATTATTTTAAATTCCTCTTGTCAGGCCGG
AATAACTCCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACG
CCGCCGGGTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAA
10 ATAAATGCTTGACTCTGTAGCGGGAA

The match between the T2 sequence and the C1/C2 sequence is

TAAATTCCTCTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTG
15 ACACGGAACAACGGCAAACACGCCCGGGTTCAGCGGGGTTCTCCTGAG
AACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCGGGAAG
GCGTATTATGCACACCCCGCGCCGCT

20 A C1/C2 short loop on chromosome 1 whose identifier is 3432 controls the
expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
as a RNA single strand that is 3'UTR to the gene btuB and has the DNA sequence

TGCGCGGTCAGAAAATTATTTTAAATTCCTCTTGTCAGGCCGGAATAACT
CCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGG
25 GTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATG
CTTGACTCTGTAGCGGGAAGGCGTATTATGCACACC...ACACCATGGGAGT
GGGTTGCAAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCGCTTACCACT
TTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAACCGTAGGGGAAC
CTGCGGTTGGATCACCTCCTTACCTTAAAGAAGCGT

30

The match between the T1 sequence and the C1/C2 sequence is

TGCGCGGTCAGAAAATTATTTTAAATTTCTTGTTCAGGCCGGAATAACT
 CCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGG
 GTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATG
 CTTGACTCTGTAGCGGGAA

5

The match between the T2 sequence and the C1/C2 sequence is

TAAATTTCTTGTTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTG
 ACACGGAACAACGGCAAACACGCCGCCGGGTCAGCGGGGTTCTCCTGAG
 AACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCGGGAAG
 GCGTATTATGCACACCCCGCGCCGCT

10

A C1/C2 short loop on chromosome 1 whose identifier is 2218 controls the
 expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
 as a RNA single strand that is 3'UTR to the gene clpB and has the DNA sequence

15

CTTGTTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGGAAC
 AACGGCAAACACGCCGCCGGGC

20

The match between the T1 sequence and the C1/C2 sequence is

CTTGTTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGGAAC
 AACGGCAAACACGCCGCCGGGC

25

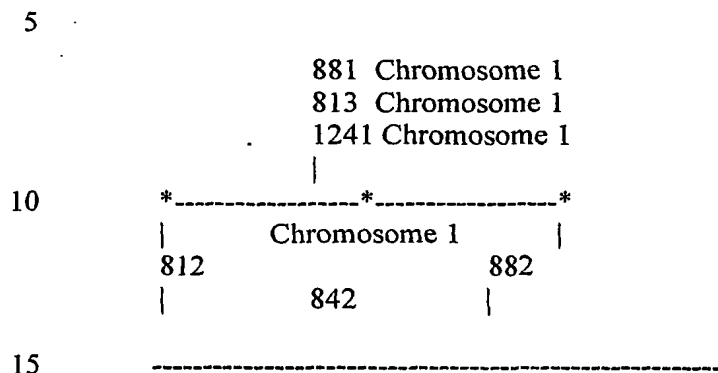
The match between the T2 sequence and the C1/C2 sequence is

CTTGTTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGGAAC
 AACGGCAAACACGCCGCCGGGTC

30

Example of an archea connectron – *H. pylori*

In this example the existence of the T1-T2 (812-882) long loop is controlled by three C1/C2 short loops (881, 813 and 1214). The T1-T2 long loop controls the expression of 54 genes on chromosome 1 in addition to one C1/C2 (843) short loop.



Connectron control elements for chromosome 1 of *H. pylori* genome

A double stranded DNA loop of length 96.385 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 812. This T1 control element has the DNA sequence

TTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 882. This T2 control element has the DNA sequence

TAGCGGAACTAAAGCATTTCATCCCAAACACTAAAGATATTTGG

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

HP0999	HP1000	HP1001	HP1002	HP1003	HP1005	HP1006
HP1008	HP1009	HPtRNA-Pro	HP1010	HP1011	HP1013	HP1015
HP1017	HP1018	HP1020	HP1021	HP1022	HP1023	HP1024

	HP1025	HP1027	HP1028	HP1030	HP1031	HP1033	HP1034
	HP1038	HP1039	HP1040	HP1041	HP1042	HP1043	HP1044
	HP1045	HP1046	HP1051	HP1052	HP1055	HP1056	HP1058
	HP1060	HP1065	HPtRNA-Ser	HP1066	HP1067	HP1069	HP1070
5	HP1074	HP1075	HP1076	HP1077	HP1078	HP1079	HP1080
	HP1081	HP1083	HP1084	HP1085	HP1088	HP1091	HP1092
	HP1093	HP1094	HP1095	HP1096			

10 This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

15 A C1/C2 short loop on chromosome 1 whose identifier is 813 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene HP0998 and has the DNA sequence

TTTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCATTTCATCCC
AAACACTAAAGATATTTGG

20 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

25 A C1/C2 short loop on chromosome 1 whose identifier is 881 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene HP1096 and has the DNA sequence

TTTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCATTTCATCCC
AAACACTAAAGATATTTGG

30 The match between the T1 sequence and the C1/C2 sequence is

TTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCA

The match between the T2 sequence and the C1/C2 sequence is

5 TAGCGGAACTAAAGCATTTCATCCCAAACACTAAAGATATTTGG

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

10 A C1/C2 short loop on chromosome 1 whose identifier is 813 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene HP0998 and has the DNA sequence

15 TTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCATTTCATCCC
AAACACTAAAGATATTTGG

A C1/C2 short loop on chromosome 1 whose identifier is 881 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene HP1096 and has the DNA sequence

20 TTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCATTTCATCCC
AAACACTAAAGATATTTGG

The match between the T1 sequence and the C1/C2 sequence is

25 TTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCA

The match between the T2 sequence and the C1/C2 sequence is

30 TAGCGGAACTAAAGCATTTCATCCCAAACACTAAAGATATTTGG

A C1/C2 short loop on chromosome 1 whose identifier is 1241 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene HP1535 and has the DNA sequence

5 TTTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCATTTCATCCC
AAACA

The match between the T1 sequence and the C1/C2 sequence is

10 TTTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCA

The match between the T2 sequence and the C1/C2 sequence is

15 TAGCGGAACTAAAGCATTTCATCCCAAACA

Example of single-celled connectron – *S. cerevisiae*

20 In this example the existence of the T1-T2 (1352-1416) long loop on chromosome 4 is controlled by one C1/C2 short loop (4213) on chromosome 10. The T1-T2 long loop controls the expression of 34 genes on chromosome 4 in addition to one C1/C2 (1356) short loop.

25 4213 Chromosome 10
|
----------*
| Chromosome 4 |
1352 1416
30 | 1356 |

Connectron control elements for chromosome 1 of *S. cerevisiae* genome

A double stranded DNA loop of length 68.908 kilo-bases on chromosome 4 is bounded on the left by a T1 sequence whose identifier is 1352. This T1 control element has the DNA sequence

5

TTATGAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 1416. This T2 control element has the DNA sequence

10

ATTAGATCTATTACATTATGGGTGGTATGTTGGAATAAAAATCAACTATCA
TCTACTAACTAGTATTTACGTTACTAGTATATTATCATATACGGTGTAGTA
AGATGACGCAAATGATGAGAAATAGTCATCTAAATTAGTGGAAGCTGAA
ACGCAAGGATTGATAATGTAATAGGATCAATGAATATTAACATATAAAAC
GATGATAATAATATTTATAGAATTGTGTAGAATTGCAGATTCCCTTTTATG
GATTCCTAAATCCTTGAGGAGAACTTCTAGTATATCTACATACCTAATATT
ATAGCCTTAATCACAATGGAATCCCAACAATTACATCAAAATCCACATTC
TCTACAGTA

15

20

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

25

YDR170W-A	YDR171W	YDR172W	YDR173C	YDR174W	YDR175C
YDR176W	YDR177W	YDR178W	YDR179C	YDR179W-A	YDR180W
YDR181C	YDR182W	YDR183W	YDR184C	YDR185C	YDR186C
YDR187C	YDR188W	YDR189W	YDR190C	YDR191W	YDR192C
YDR193W	YDR194C	YDR195W	YDR196C	YDR197W	YDR198C
YDR199W	YDR200C	YDR201W	YDR202C	YDR203W	YDR204W
YDR205W	YDR206W	YDR207C	YDR208W	YDR209C	YDR210W

30

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 4 whose identifier is 1356 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YDR170W-A and has the DNA sequence

5

AATCACAATAATCATTCTGATGATGAACTCCCTGGACACCTCCTTCTCGAT
TCAGGAGCATCACGAACCCTTATAAGATCTGCTCATCACATACTCAGC
ATCATCTAATCCTGACATAAACGTAGTTGATGCTCAAAAAAGAAATATAC
10 CAATTAACGCTATTGGTGACCTACAATTTCACTTCCAGGACAACACCAAA
ACATCAATAAAGGTATTGCACACTCCTAACATAGCCTATGACTTACTCAGT
TTGAATGAATTGGCTGCAGTAGATATCACAGCATGCTTTACCAAAAACGT
CTTAGAACG

15

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 10 whose identifier is 4213 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YJR029W and has the DNA sequence

20

ATCTATTACATTATGGGTGGTATGTTGGAATAAAAAATCCACTATCGTCTAT
CAACTAATAGTTATATTATCAATATATTATCATATACGGTGTTAAGATGAT
25 GACATAAGTTATGAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAAACGC
AAGGATTGATAATGTAATAGGATCAATGAATATAAACATATAAAACGGA
ATGAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGA
GGATTCCCTATATCCTCGAGGAGAACTTCTAGTATATTCTGTATACCTAATA
TTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAACAT

30

The match between the T1 sequence and the C1/C2 sequence is

TTATGAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAA

The match between the T2 sequence and the C1/C2 sequence is

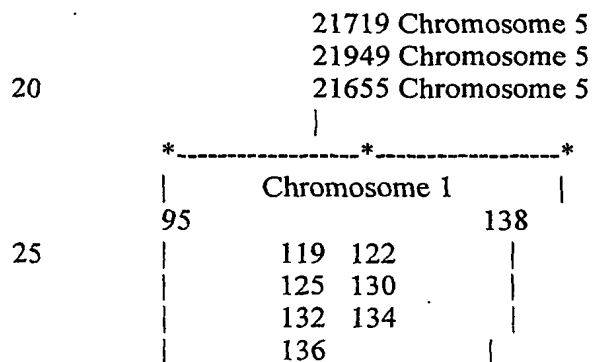
5 ATCTATTACATTATGGGTGGTATGTTGGAATAAAAATC

10

Example of a multi-celled connectron – *C. elegans*

In this example the existence of the T1-T2 (9-138) long loop on chromosome 1 is controlled by three C1/C2 short loops on chromosome 5 (21719, 21949 and 21655).

15 The T1-T2 long loop controls the expression of four genes on chromosome 1 in addition to seven C1/C2 (119, 122, 125, 130, 132, 134 and 136) short loops.



30

A double stranded DNA loop of length 41.978 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 95. This T1 control element has the DNA sequence

35

CAGCACGTTCTTAACCATGCAAAATCAGTTGAGAACTCTGCGTCTCTTCTC
CCGC

5 This double stranded DNA loop is bounded on the right by a T2 control element
whose identifier is 138. This T2 control element has the DNA sequence

ACTCTGCGTCTCTTCTCCCGCATTTTTTGTAGATCA

10 This long T1/T2 double stranded DNA loop modulates the expression of the
following genes

Y73A3A.1 Y73A3A.1 ZC123.3 ZC123.2

15 This long T1/T2 double stranded DNA loop modulates the expression of the
following C1/C2 short loops

20 A C1/C2 short loop on chromosome 1 whose identifier is 119 controls the expression
of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the gene ZC123.3 and has the
DNA sequence

TTGAGAACTCTGCGTCTCAACTCCCGCATTTTTTGTAGATCTACGTAGATC
AAACCGAAATGGGACACT

25 A C1/C2 short loop on chromosome 1 whose identifier is 122 controls the expression
of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the gene ZC123.3 and has the
DNA sequence

30 GCACGGGGTTCTGGCCTTCCTCATTGAATTTTTTCGCGCTCCATTGACAATC
GCCTGCCGGACAACGCGTGGGAAAGTCGTGTACTCCAC

A C1/C2 short loop on chromosome 1 whose identifier is 125 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ZC123.3 and has the DNA sequence

5

ACGCGCCGTAAATCTACCCCAGATATGGCCGAGCCAAAATGGCCTAGTTC
GGCAAACCTCTTTCATTTCAATTTATGAGGGAAGCCAGAA

10 A C1/C2 short loop on chromosome 1 whose identifier is 130 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ZC123.2 and has the DNA sequence

15

CTCCCGCATTTTTTGTAGATCTACGTAGATCAAACCGAAATGAGGCACTTT
CTGAATCCACGAGCTAGGCTTAAGCTTAGGCTTAAGCTTAGGCCTTTTCTC
AGGCTTAGGCTTAGGCTTA

20 A C1/C2 short loop on chromosome 1 whose identifier is 132 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ZC123.2 and has the DNA sequence

25

GCTTATGCTTGGGCTTAGGCTTAGGCGTAGGCTTAGGCTTAGGCTTAGGCT
TATGCTTAGACTTAGTCTCACTATCAGTCTTAGGCTTAGGCTTAGACTTAG
GCTTAAGCTTAGGCTTAAGCTTAGACTTAGGCTTAGGCTTAGGCTTAGGCT
TAGGCTTAGGTTTGGGCTTAGGCTTAGGCTTAACCTC

30 A C1/C2 short loop on chromosome 1 whose identifier is 134 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ZC123.2 and has the DNA sequence

TCTGCGTCTTTTCTCCCGCATTTTTTGTAGATCTACGTAGATCAAACCGAA
ATGAGGCACTTTCTGAATCCACGAGCTAGGCTTAAGCTTAGGCTTAAGCTT
AGGCCTTTTCTCAGGCTTAGGCTTAGGCTTA

5 A C1/C2 short loop on chromosome 1 whose identifier is 136 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ZC123.2 and has the DNA sequence

10 GCTTATGCTTGGGCTTAGGCTTAGGCGTAGGCTTAGGCTTAGGCTTAGGCT
TATGCTTAGACTTAGTCTCACTATCAGTCTTAGGCTTAGGCTTAGACTTAG
GCTTAAGCTTAGGCTTAAGCTTAGACTTAGGCTTAGGCTTAGGCTTAGGCT
TAGGCTTAGGTTTGGGCTTAGGCTTAGGCTTAACCTC

15 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 5 whose identifier is 21719 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
20 as a RNA single strand that is 3'UTR to the gene C39F7.5 and has the DNA sequence

ACGTTCTTAACCATGCAAAATCAGTTGAGAACTCTGCGTCTCTTCTCCCGC
ATTTTTTGTAGATC

25 The match between the T1 sequence and the C1/C2 sequence is

ACGTTCTTAACCATGCAAAATCAGTTGAGAACTCTGCGTCTCTTCTCCCGC

The match between the T2 sequence and the C1/C2 sequence is

30

ACTCTGCGTCTCTTCTCCCGCATTTTTTGTAGATC

A C1/C2 short loop on chromosome 5 whose identifier is 21949 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene F16B4.4 and has the DNA sequence

5 ACCATGCAAAATCAGTTGAGAACTCTGCGTCTCTTCTCCCGCATTTTTTGT
AGATCTACGTAGATCAAGCCGAAATGAGACACTCTGACACCACG

The match between the T1 sequence and the C1/C2 sequence is

10 ACCATGCAAAATCAGTTGAGAACTCTGCGTCTCTTCTCCCGC

The match between the T2 sequence and the C1/C2 sequence is

15 ACTCTGCGTCTCTTCTCCCGCATTTTTTGTAGATC

A C1/C2 short loop on chromosome 5 whose identifier is 21655 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene C39F7.3 and has the DNA sequence

20 AACCATGCAAAATCAGTTGAGAACTCTGCGTCTCTTCTCCCGCATTTTTTG
TAGATCTACG

The match between the T1 sequence and the C1/C2 sequence is

25 AACCATGCAAAATCAGTTGAGAACTCTGCGTCTCTTCTCCCGC

The match between the T2 sequence and the C1/C2 sequence is

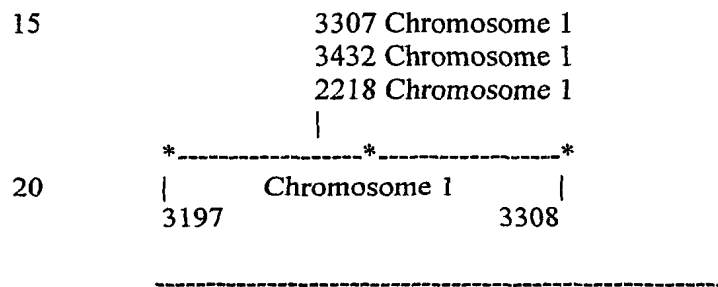
30 ACTCTGCGTCTCTTCTCCCGCATTTTTTGTAGATC

2. Many Connectrons control the expression of one set of genes in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes.

Many different C1/C2 short loops can control the existence of one T1-T2 long loop. The C1/C2 short loops can be on the same chromosome or on different chromosomes from the T1-T2 long loop. This relationship is described as “many-to-one”. This relationship exists in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes

10 Example of a many-to-one connectron in prokaryotes – E. coli

In this example the existence of the T1-T2 (3197-3308) long loop is controlled by three C1/C2 short loops (3307, 3432 and 2218).



25 A double stranded DNA loop of length 93.542 kilo-bases on chromosome 1 is
bounded on the left by a T1 sequence whose identifier is 3197. This T1 control
element has the DNA sequence

30 AAAAAATGCGCGGTCAGAAAATTATTTTAAATTTCTCTTGTCAGGCCGG
AATAACTCCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACG
CCGCCGGGTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAA
ATAAATGCTTGACTCTGTAGCGGGAA

35 This double stranded DNA loop is bounded on the right by a T2 control element
whose identifier is 3308. This T2 control element has the DNA sequence

TAAATTTCTCTTGTTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTG
 ACACGGAACAACGGCAAACACGCCGCCGGGTCAGCGGGGTTCTCCTGAG
 AACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCGGGAAG
 5 GCGTATTATGCACACCCCGCGCCGCT

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

10	rrsC	gltU	rrlC	rrfC	aspT	trpT	yifA	yifE	yifB
	ilvL	ilvG_1	ilvM	ilvE	ilvD	ilvA	ilvY	ilvC	ppiC
	b3776	rep	gppA	rhlB	trxA	rhoL	rho	rfe	wzzE
	wecB	rffH	wecD	wecE	wzxE	yifM_2	wecG	yifK	
	argX	hisR	leuT	proM	aslB	aslA	hemY	hemX	
15	hemD	cyaA	cyaY	b3808	dapF	uvrD	b3814	corA	
	yigF	yigG	rarD	yigI	pldA	recQ	yigJ	yigK	pldB
	yigL	yigM	metR	metE	ysgA	udp	yigN	ubiE	yigP
	b3836	yigU	yigW_1	rfaH	yigC	ubiB	fadA	fadB	
	pepQ	trkH	hemG						

20

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

25

A C1/C2 short loop on chromosome 1 whose identifier is 3307 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene hemG and has the DNA sequence

30

AAAAAATGCGCGGTCAGAAAATTATTTTAAATTTCTCTTGTTCAGGCCGG
 AATAACTCCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACG
 CCGCCGGGTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAA
 ATAAATGCTTGACTCTGTAGCGGGAAGGCGTATTATG...GGAGTCTGCAAC
 TCGACTCCATGAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACG

GTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGT
 GGGTTGCAAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCGCTTACCACT
 TTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAACCGTAGGGGAAC
 CTGCGGTTGGATCACCTCCTTACCTTAAAGAAGCGTTCTTTG

5

The match between the T1 sequence and the C1/C2 sequence is

AAAAAATGCGCGGTCAGAAAATTATTTTAAATTTCTCTTGTCAAGGCCGG
 AATAACTCCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACG
 10 CCGCCGGGTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAA
 ATAAATGCTTGACTCTGTAGCGGGAA

The match between the T2 sequence and the C1/C2 sequence is

15 TAAATTTCTCTTGTCAAGGCCGGAATAACTCCCTATAATGCGCCACCACTG
 ACACGGAACAACGGCAAACACGCCGCCGGGTCAGCGGGGTTCTCCTGAG
 AACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCGGGAAG
 GCGTATTATGCACACCCCGCGCCGCT

20 A C1/C2 short loop on chromosome 1 whose identifier is 3432 controls the
 expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
 as a RNA single strand that is 3'UTR to the gene btuB and has the DNA sequence

25 TGCGCGGTCAGAAAATTATTTTAAATTTCTCTTGTCAAGGCCGGAATAACT
 CCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGG
 GTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATG
 CTTGACTCTGTAGCGGGAAGGCGTATTATGCACACC...ACACCATGGGAGT
 GGGTTGCAAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCGCTTACCACT
 TTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAACCGTAGGGGAAC
 30 CTGCGGTTGGATCACCTCCTTACCTTAAAGAAGCGT

The match between the T1 sequence and the C1/C2 sequence is

5 TGC GCGGTCAGAAAATTATTTTAAATTTCTCTTGTCAGGCCGGAATAACT
CCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGG
GTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATG
CTTGACTCTGTAGCGGGAA

The match between the T2 sequence and the C1/C2 sequence is

10 TAAATTTCTCTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTG
ACACGGAACAACGGCAAACACGCCGCCGGGTCAGCGGGGTTCTCCTGAG
AACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCGGGAAG
GCGTATTATGCACACCCCGCGCCGCT

15 A C1/C2 short loop on chromosome 1 whose identifier is 2218 controls the
expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
as a RNA single strand that is 3'UTR to the gene clpB and has the DNA sequence

20 CTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGGAAC
AACGGCAAACACGCCGCCGGGC

The match between the T1 sequence and the C1/C2 sequence is

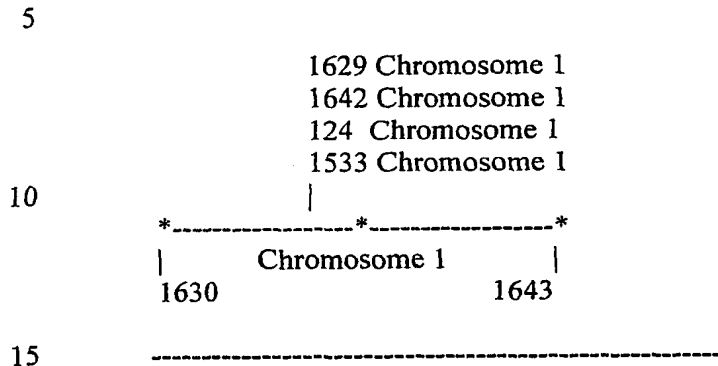
25 CTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGGAAC
AACGGCAAACACGCCGCCGGGC

The match between the T2 sequence and the C1/C2 sequence is

30 CTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGGAAC
AACGGCAAACACGCCGCCGGGC

Example of a many-to-one connectron in archea – *M. jannaschii*

In this example the existence of the T1-T2 (1630-1643) long loop is controlled by four C1/C2 short loops (1629, 1642, 124 and 1533).



A double stranded DNA loop of length 4.998 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 1630. This T1 control element has the DNA sequence

20

TTATTAATTAGTTCAAAGGATTTTATTTAATTTCTAAGGGTTTGCTGGTTT
GATTATTTAGAATATTTGAGTTTATTGAATTATTCAGATTTTAAAAATTA
AGATTAATTAGGAAAGGAAATAAGATTTCTCTAACAGACAAGTTAAATTT
TTGGATTTAAAAAGATAAAAAT

25

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 1643. This T2 control element has the DNA sequence

30

TTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTG
AATTATTCAGATTTTAAAAATTAGGATTAATTAGGCAAGTAAATAAAAAT
TTCTCTAACAAATAAGTTAAATTTTGGATTTAAAAAGATAAAAATACTCT
GTTTTATTATGGAAAGAAAGAT

35

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

MJ1597 MJ1598 MJ1599 MJ1600 MJ1601 MJ1602

5 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is 1629 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1597 and has the DNA sequence

10

ATATGTTTGAAATTTGAAAATAAGAGTATTTAGAAGTTATTAATTAGTTCA
AAGGATTTTTATTTAATTTCTAAGGGTTTGCTGGTTTGATTATTTAGAATAT
TTGAGTTTATTGAATTATTCAGATTTTTTAAAAATTA

15

The match between the T1 sequence and the C1/C2 sequence is

TTATTAATTAGTTCAAAGGATTTTTATTTAATTTCTAAGGGTTTGCTGGTTT
GATTATTTAGAATATTTGAGTTTATTGAATTATTCAGATTTTTTAAAAATTA

20

The match between the T2 sequence and the C1/C2 sequence is

GCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAATTATTCAGATTTTAA
AAAATTA

25

A C1/C2 short loop on chromosome 1 whose identifier is 1642 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1602 and has the DNA sequence

30

ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTAT
TGAATTATTCAGATTTTTTAAAAATTAGGATTAATTAGGCAAGTAAATAAA
ATTTCTCTAACAAATAAGTTAAATTTTTGGATTAAAAAAGATAAAAAATACT
CTGTTTTATTATGGAAAGAAAGAT

The match between the T1 sequence and the C1/C2 sequence is

5 GCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAATTATTCAGATTTTAA
AAAATTA

The match between the T2 sequence and the C1/C2 sequence is

10 TTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTG
AATTATTCAGATTTTTAAAAATTAGGATTAATTAGGCAAGTAAATAAAAT
TTCTCTAACAAATAAGTTAAATTTTTGGATTAAAAAGATAAAAAATACTCT
GTTTTATTATGGAAAGAAAGAT

15 A C1/C2 short loop on chromosome 1 whose identifier is 124 controls the expression
of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene MJ0112 and has the DNA sequence

20 ATTTAATTTCTAAGGGTTTGCTGGTTTGATTATTTAGAATATTTGAGTTTAT
TGAATTATTCAGATTTTTAAAAAT

The match between the T1 sequence and the C1/C2 sequence is

25 ATTTAATTTCTAAGGGTTTGCTGGTTTGATTATTTAGAATATTTGAGTTTAT
TGAATTATTCAGATTTTTAAAAAT

The match between the T2 sequence and the C1/C2 sequence is

30 GCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAATTATTCAGATTTTAA
AAAAT

A C1/C2 short loop on chromosome 1 whose identifier is 1533 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1486 and has the DNA sequence

5 TTTTATTTAATTTCTAAGGGTTTGCTGGTTTGATTATTTAGAATATTTGAG
 TTTATT

The match between the T1 sequence and the C1/C2 sequence is

10 TTTTATTTAATTTCTAAGGGTTTGCTGGTTTGATTATTTAGAATATTTGAG
 TTTATT

The match between the T2 sequence and the C1/C2 sequence is

15 GCTGGTTTGATTATTTAGAATATTTGAGTTTATT

Example of a many-to-one connectron in single-cell eukaryotes – *S. cerevisiae*

20

In this example the existence of the T1-T2 (5515-5533) long loop on chromosome 12 is controlled by seventeen C1/C2 short loops (5516, 5532, 1939, 2323, 1942, 3286, 3649, 4764, 4751, 5536, 6102, 8023, 7356, 3293, 3291, 3289 and 146).

25 5516 Chromosome 12
 5532 Chromosome 12
 1939 Chromosome 4
 2323 Chromosome 5
 1942 Chromosome 5
30 3286 Chromosome 7
 3649 Chromosome 8
 4764 Chromosome 12
 4751 Chromosome 12
 5536 Chromosome 13
35 6102 Chromosome 14

5
 8023 Chromosome 16
 7356 Chromosome 16
 3293 Chromosome 8
 3291 Chromosome 8
 3289 Chromosome 8
 146 Chromosome 2
 |
 ----------*
 10 | Chromosome 12 |
 3197 3308

15 A double stranded DNA loop of length 6.466 kilo-bases on chromosome 12 is
 bounded on the left by a T1 sequence whose identifier is 5515. This T1 control
 element has the DNA sequence

20 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
 ATTGTAAGAAATTTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
 TTCACTGTTTTGATTTAGTGTTTGTTCACGGCAGTAGCGAGAGACAAGTG
 GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
 AGGTAGTAAGTAGCTTTTGGTTG

25 This double stranded DNA loop is bounded on the right by a T2 control element
 whose identifier is 5533. This T2 control element has the DNA sequence

30 ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTTCTAGGGA
 ATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTTC
 ACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATC
 TATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACA
 TCCGGGTAAGAGACAACAGGGCT

35 This long T1/T2 double stranded DNA loop modulates the expression of the
 following genes

YLR467W

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

5 A C1/C2 short loop on chromosome 12 whose identifier is 5516 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YLR464W and has the DNA sequence

10 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTATTTAGTGTGTTGTTGCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

15 A C1/C2 short loop on chromosome 12 whose identifier is 5532 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YLR467W and has the DNA sequence

20 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTATTTAGTGTGTTGTTGCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
25 AGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

30 A C1/C2 short loop on chromosome 4 whose identifier is 1939 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed

as a RNA single strand that is 3'UTR to the gene YDR545W and has the DNA sequence

5 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTATTTAGTGTGTTGTTGCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGG

10 The match between the T1 sequence and the C1/C2 sequence is

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTATTTAGTGTGTTGTTGCACGGCAGTAGCGAGAGACAAGTG
15 GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGG

The match between the T2 sequence and the C1/C2 sequence is

20 ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGA
ATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTGTTGTTGC
ACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATC
TATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGG

25 A C1/C2 short loop on chromosome 5 whose identifier is 2323 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YER189W and has the DNA sequence

30 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTATTTAGTGTGTTGTTGCACGGCAGTAGCGAGAGACAAGTG

GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

The match between the T1 sequence and the C1/C2 sequence is

5

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTGATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTG

10

The match between the T2 sequence and the C1/C2 sequence is

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGA
ATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTC
ACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATC
TATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACA
TCCGGGTA
AGAGACAACAGGGCT

20

A C1/C2 short loop on chromosome 5 whose identifier is 1942 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YEL077C and has the DNA sequence

25

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTGATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

30

The match between the T1 sequence and the C1/C2 sequence is

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
 ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
 TTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTG
 5 GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
 AGGTAGTAAGTAGCTTTTGGTTG

The match between the T2 sequence and the C1/C2 sequence is

10 ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGA
 ATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTC
 ACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATC
 TATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACA
 TCCGGGTA
 15 AGAGACAACAGGGCT

A C1/C2 short loop on chromosome 7 whose identifier is 3286 controls the
 expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
 as a RNA single strand that is 3'UTR to the gene YGR296W and has the DNA
 20 sequence

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
 ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
 TTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTG
 25 GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
 AGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

The match between the T1 sequence and the C1/C2 sequence is

30 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
 ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
 TTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTG

GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTG

The match between the T2 sequence and the C1/C2 sequence is

5

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTCTAGGGA
ATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTGATTAGTGTTTGTTC
ACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATC
TATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACA
TCCGGGTAAGAGACAACAGGGCT

10

A C1/C2 short loop on chromosome 8 whose identifier is 3649 controls the
expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
as a RNA single strand that is 3'UTR to the gene YHR219W and has the DNA
sequence

15

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTGTAGTGTTGTTGCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

20

The match between the T1 sequence and the C1/C2 sequence is

25

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTGTAGTGTTGTTGCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTG

30

The match between the T2 sequence and the C1/C2 sequence is

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTCTAGGGA
ATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTTC
ACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATC
TATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACA
5 TCCGGGTAAGAGACAACAGGGCT

A C1/C2 short loop on chromosome 12 whose identifier is 4764 controls the
expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
as a RNA single strand that is 3'UTR to the gene YLL066C and has the DNA
10 sequence

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTATTTAGTGTTTGTTCACGGCAGTAGCGAGAGACAAGTG
15 GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

The match between the T1 sequence and the C1/C2 sequence is

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTATTTAGTGTTTGTTCACGGCAGTAGCGAGAGACAAGTG
20 GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTG

The match between the T2 sequence and the C1/C2 sequence is

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTCTAGGGA
ATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTTC
30 ACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATC
TATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACA
TCCGGGTAAGAGACAACAGGGCT

A C1/C2 short loop on chromosome 12 whose identifier is 4751 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YLL067C and has the DNA sequence

5

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTGTAGTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTG
10 GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

The match between the T1 sequence and the C1/C2 sequence is

15

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTGTAGTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTG

20

The match between the T2 sequence and the C1/C2 sequence is

25

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGA
ATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTC
ACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATC
TATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACA
TCCGGGTAAGAGACAACAGGGCT

30

A C1/C2 short loop on chromosome 13 whose identifier is 5536 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YML133C and has the DNA sequence

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTGATTAGTGTGTTGTCACGGCAGTAGCGAGAGACAAGTG
5 GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

The match between the T1 sequence and the C1/C2 sequence is

10 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTGATTAGTGTGTTGTCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTG

15

The match between the T2 sequence and the C1/C2 sequence is

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGA
ATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTGATTAGTGTGTTGTC
20 ACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATC
TATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACA
TCCGGGTAAGAGACAACAGGGCT

25 A C1/C2 short loop on chromosome 14 whose identifier is 6102 controls the
expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
as a RNA single strand that is 3'UTR to the gene YNL339C and has the DNA
sequence

30 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTGATTAGTGTGTTGTCACGGCAGTAGCGAGAGACAAGTG

GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

The match between the T1 sequence and the C1/C2 sequence is

5

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCACTGTTTTGATTAGTGTTTGTTCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
10 AGGTAGTAAGTAGCTTTTGGTTG

The match between the T2 sequence and the C1/C2 sequence is

15

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGA
ATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTGATTAGTGTTTGTTC
ACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATC
TATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACA
TCCGGGTAAGAGACAACAGGGCT

20

A C1/C2 short loop on chromosome 16 whose identifier is 8023 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YPR204W and has the DNA sequence

25

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCACTGTTTTGATTAGTGTTTGTTCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

30

The match between the T1 sequence and the C1/C2 sequence is

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTGAATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
5 AGGTAGTAAGTAGCTTTTGGTTG

The match between the T2 sequence and the C1/C2 sequence is

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGA
10 ATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTC
ACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATC
TATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACA
TCCGGGTAAGAGACAACAGGGCT

15 A C1/C2 short loop on chromosome 16 whose identifier is 7356 controls the
expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
as a RNA single strand that is 3'UTR to the gene YPL283C and has the DNA
sequence

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
20 ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTGAATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

25

The match between the T1 sequence and the C1/C2 sequence is

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
30 TTCAGTGTGTTTGAATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTG

The match between the T2 sequence and the C1/C2 sequence is

5 ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTCTAGGGA
 ATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTGATTAGTGTGTTGC
 ACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATC
 TATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACA
 TCCGGGTAAGAGACAACAGGGCT

10 A C1/C2 short loop on chromosome 8 whose identifier is 3293 controls the
 expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
 as a RNA single strand that is 3'UTR to the gene YHL050C and has the DNA
 sequence

15 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
 ATTGTAAGAAATTTTTTTTCTAGGGAATATGCGTTTT

The match between the T1 sequence and the C1/C2 sequence is

20 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
 ATTGTAAGAAATTTTTTTTCTAGGGAATATGCGTTTT

The match between the T2 sequence and the C1/C2 sequence is

25 ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTCTAGGGA
 ATATGCGTTTT

30 A C1/C2 short loop on chromosome 8 whose identifier is 3291 controls the
 expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
 as a RNA single strand that is 3'UTR to the gene YHL050C and has the DNA
 sequence

ATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGC
GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAA

The match between the T1 sequence and the C1/C2 sequence is

5

ATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGC
GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAA

The match between the T2 sequence and the C1/C2 sequence is

10

ATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGC
GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAA

A C1/C2 short loop on chromosome 2 whose identifier is 145 controls the expression
of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene YBL113C and has the DNA sequence

15

CTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
ATCCGGGTAAGAGACAACAGGCT

20

The match between the T1 sequence and the C1/C2 sequence is

CTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG

25

The match between the T2 sequence and the C1/C2 sequence is

CTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
ATCCGGGTAAGAGACAACAGGCT

30

A C1/C2 short loop on chromosome 8 whose identifier is 3289 controls the
expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed

as a RNA single strand that is 3'UTR to the gene YHL050C and has the DNA sequence

5 CTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
ATCCGGGTAAGAGACAACAGGCT

The match between the T1 sequence and the C1/C2 sequence is

10 CTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG

The match between the T2 sequence and the C1/C2 sequence is

15 CTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
ATCCGGGTAAGAGACAACAGGCT

A C1/C2 short loop on chromosome 2 whose identifier is 146 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YBL113C and has the DNA sequence

20 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAA

The match between the T1 sequence and the C1/C2 sequence is

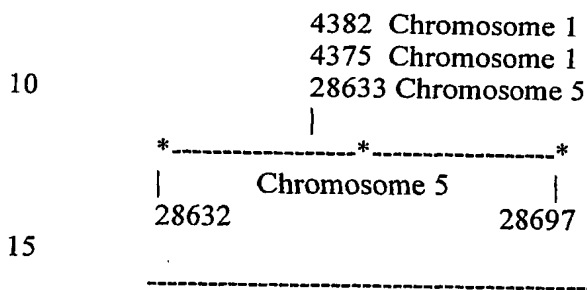
25 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAA

The match between the T2 sequence and the C1/C2 sequence is

30 ATTATGTATTGTGTAGTATAGTATATTGTAAGAAA

Example of a many-to-one connectron in multi-cell eukaryotes – *C. elegans*

In this example the existence of the T1-T2 (3197-3308) long loop on chromosome 5 is controlled by three C1/C2 short loops (4382, 4375 and 28633).



A double stranded DNA loop of length 58,451 kilo-bases on chromosome 5 is bounded on the left by a T1 sequence whose identifier is 28632. This T1 control element has the DNA sequence

GCAAAAATTGACTGAAAATTTGAATTTCCCGCAAAAATTGACTGAAAAT
TTGAATTTCCCGCCAAAATTGACTGAAAATTTGAA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 28697. This T2 control element has the DNA sequence

CAAAAATTGACTGAAAATTTGAATTTCCCTCCAAAATTGACTGAAAAT
TTGAATTTCCCGCCAAAATTGACTGAAAATTTGAATATCCCGCCAAAAA
TTGACTGAAAATTTGAATTTCCCGCCGAAAATTAAATGAAAATGGAATT
TCTCGCCGAA

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

M162.8 M162.4 M162.3 M162.6 M162.2 M162.1 M162.7

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

5

A C1/C2 short loop on chromosome 1 whose identifier is 4382 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene Y43F8B.10 and has the DNA sequence

10

ATTATAGAAAATTTAAATTTCCCTCCAAAAAATTGACTGAAAATTTGAATT
TCCCTCCAAAAAATTGACTGAAAATTTGAATTTCCCGCCAAAAAATTGACTG
AAAATTTGAATATCCCGCCAAAAAATTGACTGAAAATTTGAATTTCCCGCC
GAAAATTTAAATGAAAAATGGAATTTCTCGCCGAAAAATTCAGTAAAAATT
15 TGAATTTCTTGCCAAAAAATTGACTGAAAATTTGAATTTCTTGCCAAAAAA
GTGACTGGGAATTTGAATTTCCCTCCAAAAAATTGACTGAAATTTTGAATTT
CCCGCTAAAAGTTGACT

15

The match between the T1 sequence and the C1/C2 sequence is

20

CAAAAATTGACTGAAAATTTGAATTTCCCGC

The match between the T2 sequence and the C1/C2 sequence is

25

CAAAAATTGACTGAAAATTTGAATTTCCCTCCAAAAAATTGACTGAAAAT
TTGAATTTCCCGCCAAAAAATTGACTGAAAATTTGAATATCCCGCCAAAAA
TTGACTGAAAATTTGAATTTCCCGCCGAAAAATTAAATGAAAAATGGAATT
TCTCGCCGAA

30

A C1/C2 short loop on chromosome 1 whose identifier is 4375 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed

as a RNA single strand that is 3'UTR to the gene Y43F8B.10 and has the DNA sequence

5 ATTATAGAAAATTTAAATTTCCCTCCAAAAAATTGACTGAAAATTTGAATT
TCCCTCCAAAAAATTGACTGAAAATTTGAATTTCCCGCCAAAAAATTGACTG
AAAATTTGAATATCCCGCCAAAAAATTGACTGAAAATTTGAATTTCCCGCC
GAAAATTAAATGAAAAATGGAATTTCTCGCCGAAAAATTCAGTAAAAATT
TGAATTTCTTGCCAAAAAATTGACTGAAAATTTGAATTTCTTGCCAAAAA
GTGACTGGGAATTTGAATTTCCCTCCAAAAAATTGACTGAAAATTTGAATTT
10 CCCGCTAAAAGTTGACT

The match between the T1 sequence and the C1/C2 sequence is

15 CAAAAATTGACTGAAAATTTGAATTTCCCGC

The match between the T2 sequence and the C1/C2 sequence is

20 CAAAAAATTGACTGAAAATTTGAATTTCCCTCCAAAAAATTGACTGAAAAT
TTGAATTTCCCGCCAAAAAATTGACTGAAAATTTGAATATCCCGCCAAAAA
TTGACTGAAAATTTGAATTTCCCGCCGAAAATTAAATGAAAAATGGAATT
TCTCGCCGAA

25 A C1/C2 short loop on chromosome 5 whose identifier is 28633 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene M162.5 and has the DNA sequence

30 CAAAAATTGACTGAAAATTTGAATTTCCCGCAAAAAAATTGACTGAAAATT
TGAATTTCCCGCCAAAAAATTGACTGAAAATTTGAA

The match between the T1 sequence and the C1/C2 sequence is

CAAAAATTGACTGAAAATTTGAATTTCCCGCAAAAATTGACTGAAAATT
TGAATTTCCCGCCAAAATTGACTGAAAATTTGAA

The match between the T2 sequence and the C1/C2 sequence is

5

CAAAAATTGACTGAAAATTTGAATTTCCC

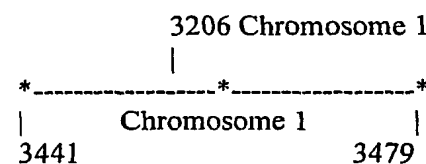
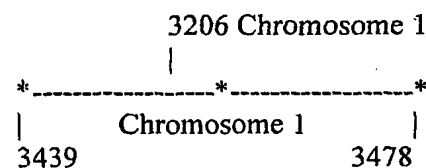
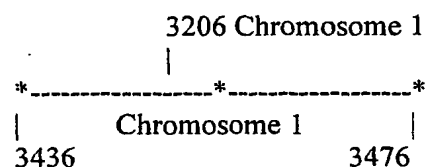
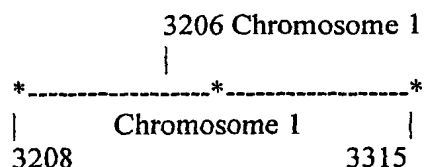
10

3. One connectron controls the expression of many sets of genes in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes.

One C1/C2 short loop can control the existence of a many T1-T2 long loops. The C1/C2 short loop can be on the same chromosome or on different chromosomes from the T1-T2 long loops. This relationship is described as "one-to-many". This relationship exists in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes.

Example of a one-to-many connectron in prokaryotes – E. coli

In this example the existence of T1-T2 (3208-3315, 3436-3476, 3439-3478 and 3441-3479) long loops are controlled by one C1/C2 short loop (3206).



A double stranded DNA loop of length 93.377 kilo-bases on chromosome 1 is
 5 bounded on the left by a T1 sequence whose identifier is 3208. This T1 control
 element has the DNA sequence

ACTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCA
 AGCTGAAAATTGAAACACTGAACAACGAAAGTTGTTTCGTGAGTCTCTCAA
 10 ATTTTCGCAACACGATGATGAATCGAAAGAAACATCTTCGGGTTGTGAGG
 TTAAGCGACTAAGCGTACACGGTGGATGCCCTGGC...AGTGTGTTTCGACA
 CACTATCATTA ACTGAATCCATAGGTTAATGAGGCGAACCGGGGGA ACTG
 AAACATCTAAGTACCCCGAGGAAAAGAAATCAACCGAGATTCCCCCAGTA
 GCGGCGAGCGAACGGGGAGCAGCCCAGAGCCTGAATCAGT

15 This double stranded DNA loop is bounded on the right by a T2 control element
 whose identifier is 3315. This T2 control element has the DNA sequence

TTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAAC
 20 GAAAGTTGTTTCGTGAGTCTCTCAAATTTTCGCAACTCTGAAGTGAAACATC
 TTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCCCTGGCA
 GTCAGAGGCGATGAAGGACGTGCTAATCTGCGATA...GGTTAATGAGGCG
 AACCGGGGGA ACTGAAACATCTAAGTACCCCGAGGAAAAGAAATCAACC
 GAGATTCCCCCAGTAGCGGCGAGCGAACGGGGAGCAGCCCAGAGCCTGA
 25 ATCAGTGTGTGTGTTAGTGGAAGCGTCTGGAAA

This long T1/T2 double stranded DNA loop modulates the expression of the
 following genes

30	rrlC	rrfC	aspT	trpT	yifA	yifE	yifB	ilvL	ilvG_1
	ilvM	ilvE	ilvD	ilvA	ilvY	ilvC	ppiC	b3776	rep
	gppA	rhlB	trxA	rhoL	rho	rfe	wzzE	wecB	rffH

wecD wecE wzxE yifM_2 wecG yifK argX hisR
 leuT proM aslB aslA hemY hemX hemD cyaA
 cyaY b3808 dapF uvrD b3814 corA yigF yigG rãrD
 yigI pldA recQ yigJ yigK pldB yigL yigM metR
 5 metE ysgA udp yigN ubiE yigP b3836 yigU
 yigW_1 rfaH yigC ubiB fadA fadB pepQ trkH
 hemG rrsA ileT

10 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is 3206 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene rrsC and has the DNA sequence

15 GTCCCCCTTCGTCTAGAGGCCAGGACACCGCCCTTTCACGGCGGTAACAG
 GGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTC
 ACCTGCCTTAATATCTCAAACTCATCTTCGGGTGATGTTTGAGATATTTG
 CTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAA...ACCGGCGATTTCG
 20 AATGGGGAAACCCAGTGTGTTTCGACACACTATCATTAAGTGAATCCATA
 GGTTAATGAGGCGAACCGGGGGAAGTGAACATCTAAGTACCCCGAGGA
 AAAGAAATCAACCGAGATTCCCCCAGTAGCGGCGAGCGAACGGGGAGCA
 GCCCAGAGCCTGAATCAGT

25 The match between the T1 sequence and the C1/C2 sequence is

ACTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCA
 AGCTGAAAATTGAAACACTGAACAACGAAAGTTGTTCGTGAGTCTCTCAA
 ATTTTCGCAACACGATGATGAATCGAAAGAAACATCTTCGGGTGAGG
 30 TTAAGCGACTAAGCGTACACGGTGGATGCCCTGGC...AGTGTGTTTCGACA
 CACTATCATTAAGTGAATCCATAGGTTAATGAGGCGAACCGGGGGAAGTGA

AAACATCTAAGTACCCCGAGGAAAAGAAATCAACCGAGATTCCCCCAGTA
GCGGCGAGCGAACGGGGAGCAGCCCAGAGCCTGAATCAGT

The match between the T2 sequence and the C1/C2 sequence is

5

TTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAAC
GAAAGTTGTTCGTGAGTCTCTCAAATTTTCGCAAC

10

A double stranded DNA loop of length 41.279 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 3436. This T1 control element has the DNA sequence

15

ACGCAACGCGTGATAAGCAATTTTCGTGTCCCCTTCGTCTAGAGGCCAG
GACACCGCCCTTTCACGGCGGTAACAGGGGTTCGAATCCCCTAGGGGACG
CCACTTGCTGGTT

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 3476. This T2 control element has the DNA sequence

20

AGTGAAAAGCAAGGCGTCTTGCGAAGCAGACTGATACGTCCCCTTCGTCT
AGAGGCCAGGACACCGCCCTTTCACGGCGGTAACAGGGGTTCGAATCCC
CTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATA

25

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

30

gltT	rrlB	rrfB	murB	coaA	b3975	tyrU	thrT	tufB
secE	nusG	rplK	rplA	rplJ	rplL	rpoB	rpoC	htrC
thiH	thiF	thiE	yjaE	yjaD	hemE	nfi	yjaG	hupA
yjaH	yjaI	hydH	purD	purH				

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

5 A C1/C2 short loop on chromosome 1 whose identifier is 3206 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene rrsC and has the DNA sequence

10 GTCCCCCTTCGTCTAGAGGCCAGGACACCGCCCTTTCACGGCGGTAACAG
GGGTTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTC
ACCTGCCTTAATATCTCAAACTCATCTTCGGGTGATGTTTGAGATATTTG
CTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAAA
GTTGTTTCGTGAGTCTCTCAAATTTTCGCAACACGATGATGAATCGAAAGA
AACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCC
15 CTGGCAGTCAGAGGCGATGAAGGACGTGCTAATCTGCGATAAGCGTCGGT
AAGGTGATATGAACCGTTATAACCGGCGATTTCCGAATGGGGAAACCCAG
TGTGTTTCGACACACTATCATTAAGTGAATCCATAGGTTAATGAGGCGAA
CCGGGGGAACTGAAACATCTAAGTACCCGAGGAAAAGAAATCAACCGA
GATTCCCCCAGTAGCGGCGAGCGAACGGGGAGCAGCCCAGAGCCTGAAT
20 CAGT

The match between the T1 sequence and the C1/C2 sequence is

25 GTCCCCCTTCGTCTAGAGGCCAGGACACCGCCCTTTCACGGCGGTAACAG
GGGTTTCGAATCCCCTAGGGGACGCCACTTGCTGGTT

The match between the T2 sequence and the C1/C2 sequence is

30 GTCCCCCTTCGTCTAGAGGCCAGGACACCGCCCTTTCACGGCGGTAACAG
GGGTTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTC
ACCTGCCTTAATA

A double stranded DNA loop of length 41.336 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 3439. This T1 control element has the DNA sequence

CCTTAATATCTCAAAACTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTT
AAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 3478. This T2 control element has the DNA sequence

GTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTG
AAACACTGAACAACGAAAGTTGTTTCGTGAGTCTCTCAAATTTT

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

rrlB	rrfB	murB	coaA	b3975	tyrU	thrT	tufB	secE
nusG	rplK	rplA	rplJ	rplL	rpoB	rpoC	htrC	thiH
thiF	thiE	yjaE	yjaD	hemE	nfi	yjaG	hupA	yjaH
yjaI	hydH	purD	purH	gltV				

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is 3206 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the generrC and has the DNA sequence

GTCCCCTTCGTCTAGAGGCCAGGACACCGCCCTTTCACGGCGGTAACAG
GGGTTCTGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTC

ACCTGCCTTAATATCTCAAAACTCATCTTCGGGTGATGTTTGAGATATTTG
 CTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAA...ACCGGCGATTTCCG
 AATGGGGGAAACCCAGTGTGTTTCGACACACTATCATTAAGTGAATCCATA
 GGTTAATGAGGCGAACCGGGGGAAGTGAACATCTAAGTACCCCGAGGA
 5 AAAGAAATCAACCGAGATTCCCCCAGTAGCGGCGAGCGAACGGGGAGCA
 GCCCAGAGCCTGAATCAGT

The match between the T1 sequence and the C1/C2 sequence is

10 CCTTAATATCTCAAAACTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTT
 AAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGA

The match between the T2 sequence and the C1/C2 sequence is

15 GTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTG
 AAACACTGAACAACGAAAGTTGTTTCGTGAGTCTCTCAAATTTT

20 A double stranded DNA loop of length 38.285 kilo-bases on chromosome 1 is
 bounded on the left by a T1 sequence whose identifier is 3441. This T1 control
 element has the DNA sequence

AATTTTCGCAACACGATGATGAATCGAAAGAAACATCTTCGGGTGTTGAG
 25 GTTAAGCGACTAAGCGTACACGGTGGATGCCCTGGCAGTCAGAGGCGATG
 AAGGACGTGCTAATCTGCGATAAGCGTCGGTAAGGTGATATGAACCGTTA
 TAACCGGCGATTTCCGAATGGGGAAACCCAGTGTGT...GATGAGAGAAGA
 TTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAA
 CAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCC
 30 GAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCC
 ATGCGAG

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 3479. This T2 control element has the DNA sequence

5 AAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGA
 TGCCCTGGCAGTCAGAGGCGATGAAGGACGTGCTAATCTGCGATAAGCGT
 CGGTAAGGTGATATGAACCGTTATAACCGGCGATTTC CGAATGGGGAAAC
 CCAGTGTGTTTCGACACACTATCATTA ACTGAATCC...CAGATTAAATCAG
 AACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGC
 GGTGGTCCCACCTGACCCCATGCCGA ACTCAGAAAGTGAAACGCCGTAGCG
 10 CCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGA ACTGCCAGGCA
 TCAAATTA

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

15 rrlB rrfB murB coaA b3975 tyrU thrT tufB secE
 nusG rplK rplA rplJ rplL rpoB rpoC htrC thiH
 thiF thiE yjaE yjaD hemE nfi yjaG hupA yjaH
 yjaI hydH purD purH gltV

20 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

25 A C1/C2 short loop on chromosome 1 whose identifier is 3206 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene rrsC and has the DNA sequence

30 GTCCCCTTCGTCTAGAGGCC CAGGACACCGCCCTTTCACGGCGGTAACAG
 GGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTC
 ACCTGCCTTAATATCTCAAAACTCATCTTCGGGTGATGTTTGAGATATTTG
 CTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAA...ACCGGCGATTTC CG
 AATGGGGAAACCCAGTGTGTTTCGACACACTATCATTA ACTGAATCCATA

GGTTAATGAGGCGAACCGGGGGAAGTGAACATCTAAGTACCCCGAGGA
 AAAGAAATCAACCGAGATTCCCCCAGTAGCGGCGAGCGAACGGGGAGCA
 GCCCAGAGCCTGAATCAGT

5 The match between the T1 sequence and the C1/C2 sequence is

AATTTTCGCAACACGATGATGAATCGAAAGAAACATCTTCGGGTTGTGAG
 GTTAAGCGACTAAGCGTACACGGTGGATGCCCTGGCAGTCAGAGGCGATG
 AAGGACGTGCTAATCTGCGATAAGCGTCGGTAAGGTGATATGAACCGTTA
 10 TAACCGGCGATTTCCGAATGGGGAAACCCAGTGTGTTTCGACACACTATC
 ATTAAGTGAATCCATAGGTTAATGAGGCGAACCGGGGGAAGTGAACATC
 TAAGTACCCCGAGGAAAAGAAATCAACCGAGATTCCCCCAGTAGCGGCG
 AGCGAACGGGGAGCAGCCCAGAGCCTGAATCAGT

15 The match between the T2 sequence and the C1/C2 sequence is

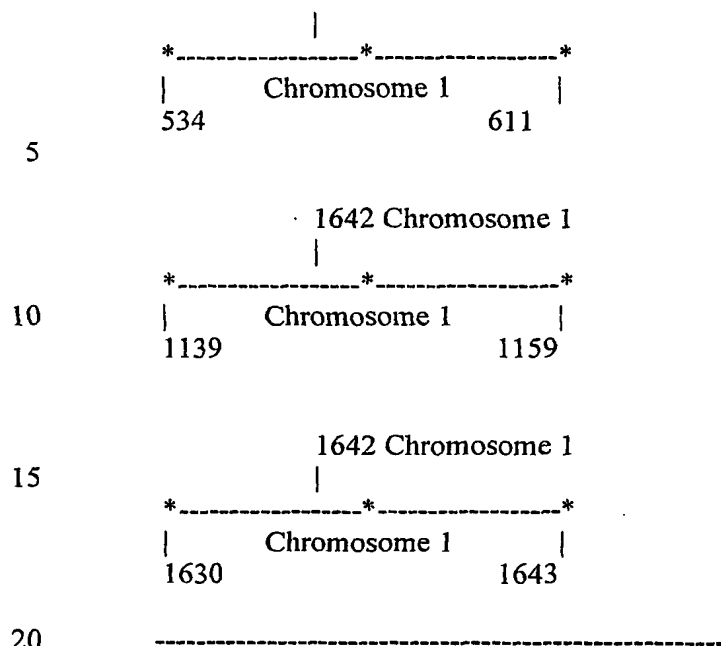
AAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGG
 TGCCCTGGCAGTCAGAGGCGATGAAGGACGTGCTAATCTGCGATAAGCGT
 CGGTAAGGTGATATGAACCGTTATAACCGGCGATTTCCGAATGGGGAAAC
 20 CCAGTGTGTTTCGACACACTATCATTAACTGAATCCATAGGTTAATGAGGC
 GAACCGGGGGAAGTGAACATCTAAGTACCCCGAGGAAAAGAAATCAAC
 CGAGATTCCCCCAGTAGCGGCGAGCGAACGGGGAGCAGCCCAGAGCCTG
 AATCAGT

25 -----

Example of a one-to-many connectron in archea – *M. jannaschii*

30 In this example the existence of T1-T2 (534-611, 1139-1159, and 1630-1643) long
 loops are controlled by one C1/C2 short loop (1642).

1642 Chromosome 1



A double stranded DNA loop of length 72.886 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 534. This T1 control element has the DNA sequence

TAAGTAAATAAAATTTCTCTAACAAATAAGTTAAATT

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 611. This T2 control element has the DNA sequence

TAAATAAAATTTCTCTAACAAATAAGTTAAATTTTGGATTAAAAAGATA
AAAATGCT

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

MJ0486	MJ0487	MJ0488	MJ0489	MJ0490	MJ0492	MJ0493
MJ0494	MJ0495	MJ0496	MJ0497	MJ0499	MJ0500	MJ0501

MJ0502	MJ0503	MJ0504	MJ0506	MJ0507	MJ0508	MJ0509
MJ0510	MJ0511	MJ0512	MJ0513	MJ0514	MJ0514	MJ0517
MJ0519	MJ0520	MJ0521	MJ0522	MJ0523	MJ0525	MJ0526
MJ0526	MJ0529	MJ0530	MJ0531	MJ0532	MJ0534	MJ0535
MJ0536	MJ0538	MJ0539	MJ0540	MJ0541	MJ0542	MJ0543
MJ0544	MJ0545	MJ0547	MJ0548	MJ0549	MJ0550	MJ0552
MJ0553	MJ0554	MJ0555	MJ0556	MJ0558	MJ0559	MJ0560
MJ0561	MJ0562	MJ0563	MJ0564			

10 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

15 A C1/C2 short loop on chromosome 1 whose identifier is 1642 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1602 and has the DNA sequence

20 ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTAT
TGAATTATTCAGATTTTAAAAATTAGGATTAATTAGGCAAGTAAATAAA
ATTTCTCTAACAAATAAGTTAAATTTTGGATTAAAAAGATAAAAAATACT
CTGTTTTATTATGGAAAGAAAGAT

The match between the T1 sequence and the C1/C2 sequence is

25 AAGTAAATAAAATTTCTCTAACAAATAAGTTAAATT

The match between the T2 sequence and the C1/C2 sequence is

30 TAAATAAAATTTCTCTAACAAATAAGTTAAATTTTGGATTAAAAAGATA
AAAAT

A double stranded DNA loop of length 14.509 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 1139. This T1 control element has the DNA sequence

5 ATTTATTAATTAGTTCAAAGGATTTTATTTAATTTCTAAGGGTTAGCTGG
TTTGATTGTTTAAAATATTTGAGTTTA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 1159. This T2 control element has the DNA sequence

10 ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTAT
TGAATTATTCAGATTTTAAAAATTA

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

15

MJ1096 MJ1097 tRNA-Arg-3 MJ1098 MJ1099 MJ1100 MJ1101
MJ1102 MJ1103 MJ1104 MJ1105 MJ1106 MJ1107 MJ1108

20 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is 1642 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1602 and has the DNA sequence

25

ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTAT
TGAATTATTCAGATTTTAAAAATTAGGATTAATTAGGCAAGTAAATAAA
ATTTCTCTAACAAATAAGTTAAATTTTGGATTAAAAAGATAAAAAATACT
30 CTGTTTTATTATGGAAAGAAAGAT

The match between the T1 sequence and the C1/C2 sequence is

ATTTAATTTCTAAGGGTTAGCTGGTTTGATT

The match between the T2 sequence and the C1/C2 sequence is

5

ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTAT
TGAATTATTCAGATTTTAAAAATTA

10

A double stranded DNA loop of length 4.998 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 1630. This T1 control element has the DNA sequence

15

TTATTAATTAGTTCAAAGGATTTTATTTAATTTCTAAGGGTTTGCTGGTTT
GATTATTTAGAATATTTGAGTTTATTGAATTATTCAGATTTTAAAAATTA
AGATTAATTAGGAAAGGAAATAAGATTTCTCTAACAGACAAGTTAAATTT
TTGGATTTAAAAAGATAAAAAT

20

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 1643. This T2 control element has the DNA sequence

25

TTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTG
AATTATTCAGATTTTAAAAATTAGGATTAATTAGGCAAGTAAATAAAAT
TTCTCTAACAAATAAGTTAAATTTTGGATTTAAAAAGATAAAAATACTCT
GTTTTATTATGGAAAGAAAGAT

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

30

MJ1597 MJ1598 MJ1599 MJ1600 MJ1601 MJ1602

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

5 A C1/C2 short loop on chromosome 1 whose identifier is 1642 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1602 and has the DNA sequence

ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTAT
TGAATTATTCAGATTTTAAAAAATTAGGATTAATTAGGCAAGTAAATAAA
10 ATTTCTCTAACAAATAAGTTAAATTTTGGATTAAAAAGATAAAAATACT
CTGTTTTATTATGGAAAGAAAGAT

The match between the T1 sequence and the C1/C2 sequence is

15 GCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAATTATTCAGATTTTAA
AAAATTA

The match between the T2 sequence and the C1/C2 sequence is

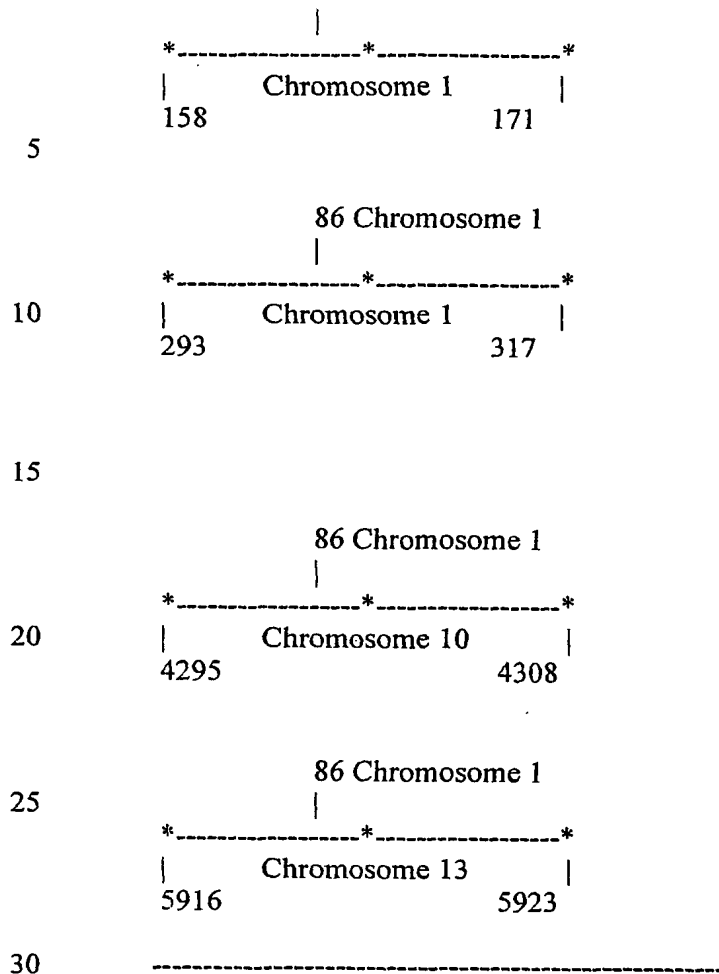
20 TTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTG
AATTATTCAGATTTTAAAAAATTAGGATTAATTAGGCAAGTAAATAAAAT
TTCTCTAACAAATAAGTTAAATTTTGGATTAAAAAGATAAAAATACTCT
GTTTTATTATGGAAAGAAAGAT

25 -----

Example of a one-to-many connectron in single-cell eukaryotes – *S. cerevisiae*

30 In this example the existence of T1-T2 (158-171, 293-317, 4295-4308 and 5916-5923) long loops are controlled by one C1/C2 short loop (86).

86 Chromosome 1



A double stranded DNA loop of length 20.391 kilo-bases on chromosome 2 is bounded on the left by a T1 sequence whose identifier is 158. This T1 control element has the DNA sequence

CCAATTGTTGGAATAAAAATCAACTATCATCTACTAACTAGTATTTACGTT
 ACTAGTATATTATCATATACGGTGTTAGAAGATGACGCAAATGATGAGAA
 ATAGTCATCTAAATTAGTGGAAGCTGAAACGCAAGGATTGATAATGTAAT
 AG

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 171. This T2 control element has the DNA sequence

ATAATTGTTGGAATAAAAAATCAACTATCATCTACTAACTAGTATTTACGTT
ACTAGTATATTATCATATACGGTGTTAGAAGATGACACAAATGATGAGAA
ATAGTCATCTAAATTAGTGGAAGCTGAAACGCAAGGATTGATAATGTAAT
5 AGGATCAATGAATATTAACATATAAAAATGATGATAATAATA

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

10 YBL107W-A TL(UAA)B1 YBL107C YBL106C YBL105C YBL104C
YBL103C YBL102W YBL101C

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

15

A C1/C2 short loop on chromosome 1 whose identifier is 86 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YAR009C and has the DNA sequence

20 ATCTATTACATTATGGGTGGTATGTTGGAATAGAAATCAACTATCATCTAC
TAACTAGTATTTACATTACTAGTATATTATCATATACGGTGTTAGAAGATG
ACGCAAATGATGAGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGCA
AGGATTGATAATGTAATAGGATCAATGAATATAAACATATAAAACGGAAT
GAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGG
25 ATTCCTATATCCTCGAGGAGAACTTCTAGTATATTCTGTATACCTAATATT
ATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAACATTCACCCAT
TTCTCAGAA

The match between the T1 sequence and the C1/C2 sequence is

30

AAATCAACTATCATCTACTAACTAGTATTTAC

The match between the T2 sequence and the C1/C2 sequence is

AAATCAACTATCATCTACTAACTAGTATTTAC

5

A double stranded DNA loop of length 38.470 kilo-bases on chromosome 2 is bounded on the left by a T1 sequence whose identifier is 293. This T1 control element has the DNA sequence

10

GAATTGTTGGAATAAAAATCCACTATCGTCTATCAACTAATAGTTATATTA
TCAATATATTATCATATACGGTGTTAAGATGATGACATAAGTTATGAGAA
GCTGTCATCGAAGTTAGAGGAAGCTGAAGTGCAAGGATTGATAATGTAAT
AGGATAATGAAACATATAAAACGGAATGAGGAATAATCGTAATATTAGT
15 ATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTTGAGGAGAAC
TTCTAGT

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 317. This T2 control element has the DNA sequence

20

AATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTC
GAGGAGAACTTCTAGTATATTCTGTA

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

25

YBL005W-B TS(AGA)B YBL004W YBL003C YBL002W YBL001C
YBR001C YBR002C YBR003W YBR004C YBR005W YBR006W
YBR007C YBR008C YBR009C YBR010W YBR011C YBR012C

30

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is 86 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YAR009C and has the DNA sequence

5

ATCTATTACATTATGGGTGGTATGTTGGAATAGAAATCAACTATCATCTAC
TAACTAGTATTTACATTACTAGTATATTATCATATACGGTGTTAGAAGATG
ACGCAAATGATGAGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGCA
AGGATTGATAATGTAATAGGATCAATGAATATAAACATATAAAACGGAAT
10 GAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGG
ATTCCTATATCCTCGAGGAGAACTTCTAGTATATTCTGTATACCTAATATT
ATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAACATTCACCCAT
TTCTCAGAA

15

The match between the T1 sequence and the C1/C2 sequence is

AAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAAT
ATAGATTCCATTTTGAGGATTCCTATATCCT

20

The match between the T2 sequence and the C1/C2 sequence is

AATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTC
GAGGAGAACTTCTAGTATATTCTGTA

25

A double stranded DNA loop of length 11.020 kilo-bases on chromosome 10 is bounded on the left by a T1 sequence whose identifier is 4295. This T1 control element has the DNA sequence

30

AAACGCAAGGATTGATAATGTAATAGGATCAATGAATATAAACATATAAA
ACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCAT
TTTGAGGATTCCTATATCCTCGAGGAGAACTTCTAGTATATTCTG

5 This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 4308. This T2 control element has the DNA sequence

GGAAGCTGAAACGCAAGGATTGATAATGTAATAGGATCAATGAATATAA
ACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAATAT
10 AGATTCCATTTTGAGGATTCCTATATCCTCGAGGAGAACTTCTAGTATATT
CTGTATACCTAATATTATAGCCTTTATCAA

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

15

YJR027W YJR029W

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

20

A C1/C2 short loop on chromosome 1 whose identifier is 87 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YAR009C and has the DNA sequence

25 ATCTATTACATTATGGGTGGTATGTTGGAATAGAAATCAACTATCATCTAC
TAACTAGTATTTACATTACTAGTATATTATCATATACGGTGTTAGAAGATG
ACGCAAATGATGAGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGCA
AGGATTGATAATGTAATAGGATCAATGAATATAAACATATAAAACGGAAT
GAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGG
30 ATTCCTATATCCTCGAGGAGAACTTCTAGTATATTCTGTATACCTAATATT
ATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAACATTCACCCAT
TTCTCA

5

A double stranded DNA loop of length 5.462 kilo-bases on chromosome 13 is bounded on the left by a T1 sequence whose identifier is 5916. This T1 control element has the DNA sequence

10

AAGCTGAAGTGCAAGGATTGATAATGTAATAGGATAATGAAACATATAA
AACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCA
TTTTGAGGATTCCTATATCCTCGAGGAGAACTTCTAGTATATTCTGTA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 5923. This T2 control element has the DNA sequence

15

TAATAGGATAATGAAACATATAAAACGGAATGAGGAATAATCGTAATATT
AGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAG
AACTTCTAGTATATTCTGTATACCTAATATTATAGCCTTTATCAA

20

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

YML045W

25

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

30

A C1/C2 short loop on chromosome 1 whose identifier is 87 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YAR009C and has the DNA sequence

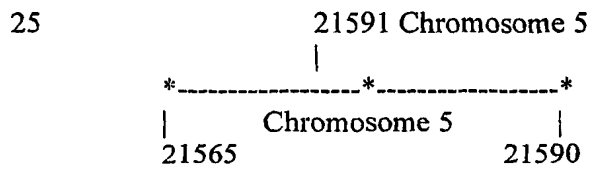
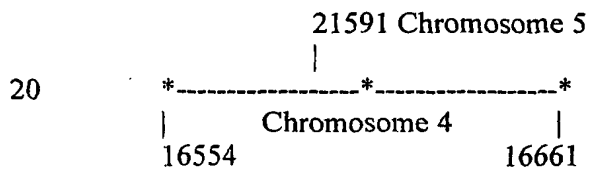
ATCTATTACATTATGGGTGGTATGTTGGAATAGAAATCAACTATCATCTAC
TAACTAGTATTTACATTACTAGTATATTATCATATACGGTGTTAGAAGATG

ACGCAAATGATGAGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGCA
 AGGATTGATAATGTAATAGGATCAATGAATATAAACATATAAAACGGAAT
 GAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGG
 ATTCCTATATCCTCGAGGAGAACTTCTAGTATATTCTGTATACCTAATATT
 5 ATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAACATTCACCCAT
 TTCTCA

10 Example of a one-to-many connectron in multi-cell eukaryotes – *C. elegans*

In this example the existence of T1-T2 (16554-16661 and 21565-21590) long loops
 are controlled by one C1/C2 short loop (21591).

15



35 A double stranded DNA loop of length 50.159 kilo-bases on chromosome 4 is
 bounded on the left by a T1 sequence whose identifier is 16554. This T1 control
 element has the DNA sequence

TGCCTGAAAAAATTGGCTCCGAGTTAGGACACTTGGGGTGGTCAAAAAAT
TTTGTGACTATTGTCAAATGAAAGATCATAGTTGATAACATAAATTCCCAA
AGTTTCATAAAAATCGATACGCAGCGAACAAAGTTATCAATT

- 5 This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 16661. This T2 control element has the DNA sequence

CACTTGGGGTGGTCAAAAAATTTTGTGATTATTGTCAAATGAAAGATCAT
GGTTGATAACATAAATTCCCAAAGTTTCATAAAAATCGATACGCAGCGAA
10 CAAAGTTATGATTTTTGACCCGGAAGTTATTTGGAGACCTA

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

- 15 C23H5.7 C23H5.8a C23H5.3 C23H5.2 C23H5.9 C23H5.1

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

- 20 A C1/C2 short loop on chromosome 5 whose identifier is 21591 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene F25A2.1 and has the DNA sequence

- 25 TATTGTCAAATGAAAGATCATGGTTGATAACATAAATTCCCACAATTCAT
AAAAATCGATACGCAGCGAACAAAGTTATGATTTTTGACCCGGAAGTTAT
TTGGAGACCTAATATT

The match between the T1 sequence and the C1/C2 sequence is

- 30 TTTCATAAAAATCGATACGCAGCGAACAAAGTTAT

The match between the T2 sequence and the C1/C2 sequence is

TATTGTCAAATGAAAGATCATGGTTGATAACATAAAATTCCCA

5

A double stranded DNA loop of length 18.142 kilo-bases on chromosome 5 is bounded on the left by a T1 sequence whose identifier is 21565. This T1 control element has the DNA sequence

10

CTCCGAGTTAGGACACTTGGGGTGGACAAAAAATTTTGTGACTATTGTCA
AATGAAAGATCATGGTTGATAA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 21590. This T2 control element has the DNA sequence

15

TATTGTCAAATGAAAGATCATGGTTGATAACATAAAATTTCCACAATTTTCAT
AAAAATCGATACGCAGCGAACAAAGTTATGATTTTTGACCCGGAAGTTAT
TTGGAGACCTAATA

20

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

T21H3.2 T21H3.1 F25A2.1

25

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 5 whose identifier is 21591 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene F25A2.1 and has the DNA sequence

30

TATTGTCAAATGAAAGATCATGGTTGATAACATAAAATTCCCACAATTCAT
AAAAATCGATACGCAGCGAACAAAGTTATGATTTTGGACCCGGAAGTTAT
TTGGAGACCTAATATT

5 The match between the T1 sequence and the C1/C2 sequence is

TATTGTCAAATGAAAGATCATGGTTGATAA

The match between the T2 sequence and the C1/C2 sequence is

10

TATTGTCAAATGAAAGATCATGGTTGATAACATAAAATTCCCACAATTCAT
AAAAATCGATACGCAGCGAACAAAGTTATGATTTTGGACCCGGAAGTTAT
TTGGAGACCTAATA

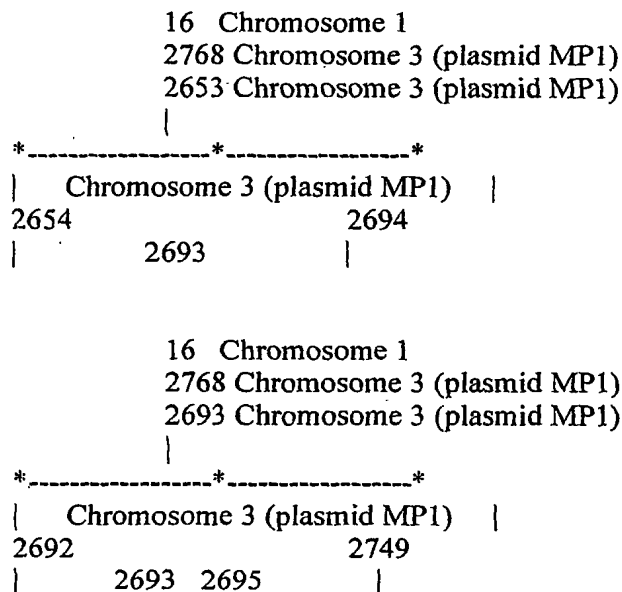
15

4. Connectrons occur between prokaryotes and their plasmids.

Connectron relationships exist between prokaryotes and their plasmids. These connectrons implement a control mechanism between the two genomes that makes it possible for them to form a symbiotic relationship. In the case of *D. radiodurans* the relationship is not symmetric. The *D. radiodurans* genome sends C1/C2 short loops to the MP1 plasmid.

Example of a prokaryote/plasmid connectron – *D. radiodurans*

In this example the existence of T1-T2 (2654-2694 and 2692-2749) long loops in chromosome 3 that is the plasmid MP1 are controlled by one C1/C2 short loop (16) in chromosome 1.



A double stranded DNA loop of length 46.903 kilo-bases on chromosome 3 (plasmid MP1) is bounded on the left by a T1 sequence whose identifier is 2654. This T1 control element has the DNA sequence

CAGCGTTTTTCTCGCTGTTCTGGACGGCTGAACGCCCTGAATCTCTCCCG
 GTATGCAGCCTGCTCGGAGAGTACGATTCGTCGTTGGCTGCACCGAAGTG
 ACGATGGGGCCATTCCGTGGGGCGCGTTACACCAGGCGACTGTCAGTACA
 5 GCAATCGAGAGTGGGCTGATCAGCCCACTGTGCGTTCTGGCCATCGACGC
 CTCTTTTCACCGCAAAGCCGGTCAGCACACCGCACACCTCGGCTCGTTCTG
 GAATGGCTGTGCCGCGCGGACC

10 This double stranded DNA loop is bounded on the right by a T2 control element
 whose identifier is 2694. This T2 control element has the DNA sequence

GCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCGGAGAGTACGA
 TTCGTCGTTGGCTGCACCGAAGTGACGATGGGGCCATTCCGTGGGGCGCG
 TTACACCAGGCGACTGTCAGTACAGCAATCGAGAGTGGGCTGATCAGCCC
 15 ACTGTGCGTTCTGGCCATCGACGCCTCTTTTCACCGCAAAGCCGGTCAGCA
 CACCGCACACCTCGGCTCGTTCTGGAATGGCTGTGCCGCGCGGACCGAAC
 GCGGAATCGAGCAATCCTGTTGT

20 This long T1/T2 double stranded DNA loop modulates the expression of the
 following genes

DRB0020	DRB0021	DRB0022	DRB0023	DRB0024	DRB0025
DRB0027	DRB0030	DRB0032	DRB0033	DRB0034	DRB0035
DRB0037	DRB0038	DRB0039	DRB0041	DRB0042	DRB0043
25 DRB0044	DRB0045	DRB0047	DRB0051	DRB0052	DRB0054
DRB0055	DRB0057				

30 This long T1/T2 double stranded DNA loop modulates the expression of the
 following C1/C2 short loops

A C1/C2 short loop on chromosome 3 (plasmid MP1) whose identifier is 2693
 controls the expression of the genes of one or more other T1/T2 long loops. This

C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene DRB0057 and has the DNA sequence

5 CTGATGGCCATCCTACAGTACGTTCTCAGCGCGGTCCCGCTGCGCAAGAC
GCAGCGGAATTTCTGACCGTGCTGCTCAGCGTTTTTCTCGCTGTTCTCTGG
AC

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

10

A C1/C2 short loop on chromosome 1 whose identifier is 16 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene DR0009 and has the DNA sequence

15 GCTGTGAAATCACCGCTTCCAATGGGTCTGATGGCCATCCTACAGTACGTT
CTCAGCGCGGTCCCGCTGCGCAAGACGCAGCGGAATTTCTGACCGTGCT
GCTCAGCGTTTTTCTCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTC
CCGGTATGCAGCCTGCTCGGAGAGTACGATTCGT

20

The match between the T1 sequence and the C1/C2 sequence is

25 CAGCGTTTTTCTCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTCCCG
GTATGCAGCCTGCTCGGAGAGTACGATTCGTCGTTGGCTGCACCGAAGTG
ACGATGGGGCCATTCCGTGGGGCGCGTTACACCAGGCGACTGTCAGTACA
GCAATCGAGAGTGGGCTGATCAGCCCACTGTGCGTTCTGGCCATCGACGC
CTCTTTTCACCGCAAAGCCGGTCAGCACACCGCACACCTCGGCTCGTTCTG
GAATGGCTGTGCCGCGCGGACC

30

The match between the T2 sequence and the C1/C2 sequence is

GCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCGGAGAGTACGA
TTCGTCGTTGGCTGCACCGAAGTGACGATGGGGCCATTCCGTGGGGCGCG
TTACACCAGGCGACTGTCAGTACAGCAATCGAGAGTGGGCTGATCAGCCC
ACTGTGCGTTCTGGCCATCGACGCCTCTTTTCACCGCAAAGCCGGTCAGCA
5 CACCGCACACCTCGGCTCGTTCTGGAATGGCTGTGCCGCGCGGACCGAAC
GCGGAATCGAGCAATCCTGTTGT

A C1/C2 short loop on chromosome 3 (plasmid MP1) whose identifier is 2768
controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is
10 expressed as a RNA single strand that is 3'UTR to the gene DRB0133 and has the
DNA sequence

GCTGTGAAATCACCGCTTCCAATGGGTCTGATGGCCATCCTACAGTACGTT
CTCAGCGCGGTCCCGCTGCGCAAGACGCAGCGGAATTTCTGACCGTGCT
15 GCTCAGCGTTTTTCTCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTC
CCGGTATGCAGCCTGCTCGGAGAGTACGATTCGT

The match between the T1 sequence and the C1/C2 sequence is

20 CAGCGTTTTTCTCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTCCCG
GTATGCAGCCTGCTCGGAGAGTACGATTCGTCGTTGGCTGCACCGAAGTG
ACGATGGGGCCATTCCGTGGGGCGCGTTACACCAGGCGACTGTCAGTACA
GCAATCGAGAGTGGGCTGATCAGCCCACTGTGCGTTCTGGCCATCGACGC
25 CTCTTTTCACCGCAAAGCCGGTCAGCACACCGCACACCTCGGCTCGTTCTG
GAATGGCTGTGCCGCGCGGACC

The match between the T2 sequence and the C1/C2 sequence is

30 GCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCGGAGAGTACGA
TTCGTCGTTGGCTGCACCGAAGTGACGATGGGGCCATTCCGTGGGGCGCG
TTACACCAGGCGACTGTCAGTACAGCAATCGAGAGTGGGCTGATCAGCCC

ACTGTGCGTTCTGGCCATCGACGCCTCTTTTCACCGCAAAGCCGGTCAGCA
CACCGCACACCTCGGCTCGTTCTGGAATGGCTGTGCCGCGCGGACCGAAC
GCGGAATCGAGCAATCCTGTTGT

5 A C1/C2 short loop on chromosome 3 (plasmid MP1) whose identifier is 2653 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene DRB0017 and has the DNA sequence

10 CGGTCCCGCTGCGCAAGACGCAGCGGAATTTCTGACCGTGCTGCTCAGC
GTTTTTCTCGCTGTTCTGGACGGCTGAACGCCCTGAATCTCTCCCGGTAT
GCAGCCTGCTCGGAGAGTACGATTCGTCGTTGGCTGCACCGAAGTGACGA
TGGGGCCATTCCGTGGGGCGCGTTACACCAGGCGA

15 The match between the T1 sequence and the C1/C2 sequence is

CAGCGTTTTTCTCGCTGTTCTGGACGGCTGAACGCCCTGAATCTCTCCCG
GTATGCAGCCTGCTCGGAGAGTACGATTCGTCGTTGGCTGCACCGAAGTG
ACGATGGGGCCATTCCGTGGGGCGCGTTACACCAGGCGACTGTCAGTACA
20 GCAATCGAGAGTGGGCTGATCAGCCCACTGTGCGTTCTGGCCATCGACGC
CTCTTTTCACCGCAAAGCCGGTCAGCACACCGCACACCTCGGCTCGTTCTG
GAATGGCTGTGCCGCGCGGACC

The match between the T2 sequence and the C1/C2 sequence is

25 GCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCGGAGAGTACGA
TTCGTCGTTGGCTGCACCGAAGTGACGATGGGGCCATTCCGTGGGGCGCG
TTACACCAGGCGACTGTCAGTACAGCAATCGAGAGTGGGCTGATCAGCCC
ACTGTGCGTTCTGGCCATCGACGCCTCTTTTCACCGCAAAGCCGGTCAGCA
30 CACCGCACACCTCGGCTCGTTCTGGAATGGCTGTGCCGCGCGGACCGAAC
GCGGAATCGAGCAATCCTGTTGT

A double stranded DNA loop of length 68.612 kilo-bases on chromosome 3 (plasmid MP1) is bounded on the left by a T1 sequence whose identifier is 2692. This T1 control element has the DNA sequence

CTGATGGCCATCCTACAGTACGTTCTCAGCGCGGTCCCGCTGCGCAAGAC
GCAGCGGAATTTCTGACCGTGCTGCTCAGCGTTTTTCTCGCTGTTCTCTGG
AC

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 2749. This T2 control element has the DNA sequence

AGCGCGGTCCCGCTGCGCAAGACGCAGCGGAATTTCTGACCGTGCTGCT
CAGCGTTTTTCTCGCTGTTCTCTGGACGGCTGAACGCCCTGAATCTCTCCCG
GT

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

DRB0059	DRB0060	DRB0061	DRB0062	DRB0064	DRB0065
DRB0066	DRB0067	DRB0068	DRB0069	DRB0070	DRB0072
DRB0073	DRB0074	DRB0076	DRB0077	DRB0079	DRB0080
DRB0081	DRB0083	DRB0085	DRB0086	DRB0087	DRB0088
DRB0089	DRB0090	DRB0092	DRB0093	DRB0094	DRB0096
DRB0097	DRB0098	DRB0102	DRB0103	DRB0104	DRB0105
DRB0106	DRB0107	DRB0111	DRB0112		

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 3 (plasmid MP1) whose identifier is 2693 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene DRB0057 and has the DNA sequence

5

CTGATGGCCATCCTACAGTACGTTCTCAGCGCGGTCCCGCTGCGCAAGAC
GCAGCGGAATTCCTGACCGTGCTGCTCAGCGTTTTTCTCGCTGTTCTGG
AC

10

A C1/C2 short loop on chromosome 3 (plasmid MP1) whose identifier is 2695 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene DRB0057 and has the DNA sequence

15

GCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCGGAGAGTACGA
TTCGTCGTTGGCTGCACCGAAGTGACGATGGGGCCATTCCGTGGGGCGCG
TTACACCAGGCGACTGTCAGTACAGCAATCGAGAGTGGGCTGATCAGCCC
ACTGTGCGTTCTGGCCATCGACGCCTCTTTTCACCGCAAAGCCGGTCAGCA
CACCGCACACCTCGGCTCGTTCTGGAATGGCTGTGCCGCGCGGACCGAAC
20 GCGGAATCGAGCAATCCTGTTGT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

25

A C1/C2 short loop on chromosome 1 whose identifier is 16 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene DR0009 and has the DNA sequence

30

GCTGTGAAATCACCGCTTCCAATGGGTCTGATGGCCATCCTACAGTACGTT
CTCAGCGCGGTCCCGCTGCGCAAGACGCAGCGGAATTCCTGACCGTGCT
GCTCAGCGTTTTTCTCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTC
CCGGTATGCAGCCTGCTCGGAGAGTACGATTCGT

The match between the T1 sequence and the C1/C2 sequence is

5 CTGATGGCCATCCTACAGTACGTTCTCAGCGCGGTCCCGCTGCGCAAGAC
GCAGCGGAATTTCTGACCGTGCTGCTCAGCGTTTTTCTCGCTGTTCTCTGG
AC

The match between the T2 sequence and the C1/C2 sequence is

10 AGCGCGGTCCCGCTGCGCAAGACGCAGCGGAATTTCTGACCGTGCTGCT
CAGCGTTTTTCTCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTCCCG
GT

15 A C1/C2 short loop on chromosome 3 (plasmid MP1) whose identifier is 2768
controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the gene DRB0133 and has the
DNA sequence

20 GCTGTGAAATCACCGCTTCCAATGGGTCTGATGGCCATCCTACAGTACGTT
CTCAGCGCGGTCCCGCTGCGCAAGACGCAGCGGAATTTCTGACCGTGCT
GCTCAGCGTTTTTCTCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTC
CCGGTATGCAGCCTGCTCGGAGAGTACGATTCGT...CGGACCGAACGCGGA
ATCGAGCAATCCTGTTGTGCCCTCATTGATGTCCAGCACCGGCAGGCCTTG
ACGGTCGATGTCCGTCAGACCCTGACCGGGTCTGAGGCTCCAACCTCGTCT
25 GGAACAG

The match between the T1 sequence and the C1/C2 sequence is

30 CTGATGGCCATCCTACAGTACGTTCTCAGCGCGGTCCCGCTGCGCAAGAC
GCAGCGGAATTTCTGACCGTGCTGCTCAGCGTTTTTCTCGCTGTTCTCTGG
AC

The match between the T2 sequence and the C1/C2 sequence is

AGCGCGGTCCCGCTGCGCAAGACGCAGCGGAATTCCTGACCGTGCTGCT
CAGCGTTTTTCTCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTCCCG
GT

A C1/C2 short loop on chromosome 3 (plasmid MP1) whose identifier is 2693 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene DRB0057 and has the DNA sequence

CTGATGGCCATCCTACAGTACGTTCTCAGCGCGGTCCCGCTGCGCAAGAC
GCAGCGGAATTCCTGACCGTGCTGCTCAGCGTTTTTCTCGCTGTTCTGG
AC

The match between the T1 sequence and the C1/C2 sequence is

CTGATGGCCATCCTACAGTACGTTCTCAGCGCGGTCCCGCTGCGCAAGAC
GCAGCGGAATTCCTGACCGTGCTGCTCAGCGTTTTTCTCGCTGTTCTGG
AC

The match between the T2 sequence and the C1/C2 sequence is

AGCGCGGTCCCGCTGCGCAAGACGCAGCGGAATTCCTGACCGTGCTGCT
CAGCGTTTTTCTCGCTGTTCTGGAC

A C1/C2 short loop on chromosome 3 (plasmid MP1) whose identifier is 2653 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene DRB0017 and has the DNA sequence

CGGTCCCGCTGCGCAAGACGCAGCGGAATTCCTGACCGTGCTGCTCAGC
GTTTTTCTCGCTGTTTCCTGGACGGCTGAACGCCCTGAATCTCTCCCGGTAT
GCAGCCTGCTCGGAGAGTACGATTCGTCGTTGGCTGCACCGAAGTGACGA
TGGGGCCATTCCGTGGGGCGCGTTACACCAGGCGA

5

The match between the T1 sequence and the C1/C2 sequence is

CGGTCCCGCTGCGCAAGACGCAGCGGAATTCCTGACCGTGCTGCTCAGC
GTTTTTCTCGCTGTTTCCTGGAC

10

The match between the T2 sequence and the C1/C2 sequence is

CGGTCCCGCTGCGCAAGACGCAGCGGAATTCCTGACCGTGCTGCTCAGC
GTTTTTCTCGCTGTTTCCTGGACGGCTGAACGCCCTGAATCTCTCCCGGT

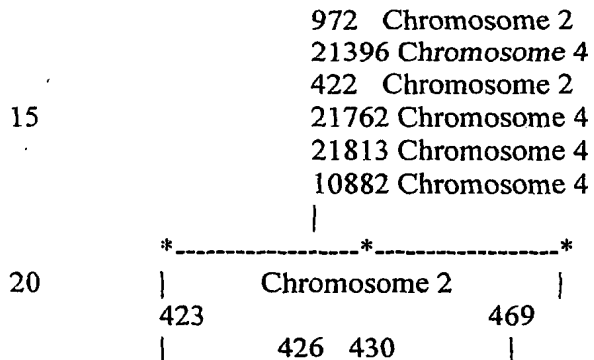
15

5. Connectrons occur in plants and higher animals

Connectron relationships exist in plant and higher animals.

5 Example of a plant connectron – *A. thaliana*

In this example the existence of the T1-T2 (423-469) long loop is controlled by six C1/C2 short loops (972, 21396, 422, 21762, 21813 and 10882). The T1-T2 long loop controls the expression of six genes on chromosome 2 in addition to two C1/C2 (426 and 430) short loops.



A double stranded DNA loop of length 42.285 kilo-bases on chromosome 2 is bounded on the left by a T1 sequence whose identifier is 423. This T1 control element has the DNA sequence

TATCTCTTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTA
ATTAAAAAACGAAATA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 469. This T2 control element has the DNA sequence

TACTAATTTAATTAATTAATAATTTAATTAAAAACGAAATACATTATTAATT
TTCAAAAATAATAACC

5 This long T1/T2 double stranded DNA loop modulates the expression of the
following genes

At2g02070 At2g02080 At2g02090 At2g02100 At2g02120 At2g02130

10 This long T1/T2 double stranded DNA loop modulates the expression of the
following C1/C2 short loops

15 A C1/C2 short loop on chromosome 2 whose identifier is 426 controls the expression
of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the gene At2g02060 and has the
DNA sequence

TTCCAAAAATAATAACCAATCAAAATCAACATATAAGATTTGATATCTAA
ATTTT

20 A C1/C2 short loop on chromosome 2 whose identifier is 430 controls the expression
of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the gene At2g02060 and has the
DNA sequence

25 TTGCGGAAAAATAATATCATCATTATAAAAAATAATTAGAGTTTTTCGC
ATAT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2
short loops.

30

A C1/C2 short loop on chromosome 2 whose identifier is 972 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene At2g04240 and has the DNA sequence

5 GTATGCCATTAGAAATAAAATTTTAAAAGTAAATTAATTCATCTCTTTAAA
AATTA AAAAGTCAAATACTAATTTAATTAATTAATTAATTAATTA AAAAACG
AAATACATTATTAATTT

The match between the T1 sequence and the C1/C2 sequence is

10 ATTAAAAAGTCAAATACTAATTTAATTAATTAATTAATTAATTA AAAAACG
AATA

The match between the T2 sequence and the C1/C2 sequence is

15 TACTAATTTAATTAATTAATTAATTAATTA AAAAACGAAATACATTATTAATT
T

A C1/C2 short loop on chromosome 4 whose identifier is 21396 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene AT4g15300 and has the DNA sequence

20 TGCCATTAGAAATAAAATTTTAAAGAGTAAATTAATTTATCTCTTTAAGGA
25 TTA AAAAGTCAAATACTAATTTAATTAATTAATTAATTAATTA AAAAACGAA
ATACATTATTAATTTCCAAAA

The match between the T1 sequence and the C1/C2 sequence is

30 TATCTCTTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAATTAATTTA
ATTAAAAAACGAAATA

The match between the T2 sequence and the C1/C2 sequence is

TACTAATTTAATTAATTAATAATTTAATTAAAAACGAAATACATTATTAATT
T

5

A C1/C2 short loop on chromosome 2 whose identifier is 422 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene At2g02060 and has the DNA sequence

10

TAACCTTAATTTTTGTAAGTAATTATATAGGTATGCCATTAGAAATAAAAT
TTTAAAGAGTAAATTAATTTATCTCTTTAAGGATTAAAAAGTCAAATACTA
ATTTAATTAATTAATAATTTAATTAATAAAAAACGAAATA

The match between the T1 sequence and the C1/C2 sequence is

15

TATCTCTTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAATTTA
ATTAAAAAACGAAATA

The match between the T2 sequence and the C1/C2 sequence is

20

TACTAATTTAATTAATTAATAATTTAATTAAAAACGAAATA

A C1/C2 short loop on chromosome 4 whose identifier is 21762 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene AT4g17510 and has the DNA sequence

25

TTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAATTTAATTTAA
AACGAAATACATT

30

The match between the T1 sequence and the C1/C2 sequence is

TTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTAATTAAA
AAACGAAATA

The match between the T2 sequence and the C1/C2 sequence is

5

TACTAATTTAATTAATTAAATTTAATTAAAAACGAAATACATT

A C1/C2 short loop on chromosome 4 whose identifier is 21813 controls the
expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
as a RNA single strand that is 3'UTR to the gene AT4g17680 and has the DNA
sequence

10

TTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTAATTAAA
AAACGAAATACATT

15

The match between the T1 sequence and the C1/C2 sequence is

TTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTAATTAAA
AAACGAAATA

20

The match between the T2 sequence and the C1/C2 sequence is

TACTAATTTAATTAATTAAATTTAATTAAAAACGAAATACATT

25

A C1/C2 short loop on chromosome 2 whose identifier is 10882 controls the
expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
as a RNA single strand that is 3'UTR to the gene At2g26540 and has the DNA
sequence

30

TATCTCTTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTA
ATTAA

The match between the T1 sequence and the C1/C2 sequence is

TATCTCTTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTA
ATTAA

5

The match between the T2 sequence and the C1/C2 sequence is

TACTAATTTAATTAATTAAATTTAATTAA

10

Example of a animal connectron – D. megalomaster

15

A double stranded DNA loop of length 88.159 kilo-bases on chromosome 4 is
bounded on the left by a T1 sequence whose identifier is 3340. This T1 control
element has the DNA sequence

20

ACCTAAAAGAAGTACCGTTTTTTACTCCTAATTACCAATTCTAACCATCCA
TATCACTTTTTGACGGACTCCGTGAAAATAATTTTTGGCCAAATTTTCGCA
TTTTTTGTAAGGGGTAACATCATAAAAATT

25

This double stranded DNA loop is bounded on the right by a T2 control element
whose identifier is 3372. This T2 control element has the DNA sequence

AAAAAAGTACCGCGTTTTACTCCTAATTACCAATTCTAACCATCCATATCA
CTTTTTGACGGACTCCGTGAAAATAATTTTTGGCCAAATTTTCGCATTTTTT
GTAAGGGGTAACATCATCAAAATTTGCGAAAAA

30

This long T1/T2 double stranded DNA loop modulates the expression of the
following genes

[Some of the following gene names have not been determined.]

- - - - -
 - CG11207 - CG2186 CG2157
 - Ork1 - - -
 5 -

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

10 A C1/C2 short loop on chromosome 4 whose identifier is 3362 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene XXX and has the DNA sequence

15 AAAAAAGTACCGCGTTTTACTCCTAATTACCAATTCTAACCATCCATATCA
 CTTTTTGACGGACTCCGTTAAAATAATTTTTGACCAAATTTTCGCATTTTTT
 GTAATCAAATTTGCAAAAAATTGAAAAAAC

20 A C1/C2 short loop on chromosome 4 whose identifier is 3364 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene XXX and has the DNA sequence

25 CAAAATTTGAATGCAAATCGATTGGGAATCAAAAAACAACTCAACGAG
 GTATGACATTCCATATTTGGGCCATTATTTCCAA

30 A C1/C2 short loop on chromosome 4 whose identifier is 3366 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene XXX and has the DNA sequence

TTTTTTCACAAAAATTAGGAAAATGATTTTGGGTAAAAAAATGAATATTT
AAGTTGGGTTTT

5 A C1/C2 short loop on chromosome 4 whose identifier is 3369 controls the
expression of the genes of one or more other T1/T2 long loops. This C1/C2 short
loop is expressed as a RNA single strand that is 3'UTR to the gene XXX and has the
DNA sequence

10 AAATCGATTGGGAATCAAAAAACAAACCTCAACGAGGTATGACATTCCAT
ATCTGGGCCATTATTTCCAATCTTTTGATCAAAATAC

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2
short loops.

15 A C1/C2 short loop on chromosome 4 whose identifier is 3373 controls the
expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
as a RNA single strand that is 3'UTR to the gene XXX and has the DNA sequence

20 AAAAAAGTACCGCGTTTTACTCCTAATTACCAATTCTAACCATCCATATCA
CTTTTGTACGGACTCCGTGAAAATAATTTTGGCCAAATTTTCGCATTTTTT
GTAAGGGGTAACATCATCAAAATTTGCGAAAAA

The match between the T1 sequence and the C1/C2 sequence is

25 TTTTACTCCTAATTACCAATTCTAACCATCCATATCACTTTTTGACGGACTC
CGTGAAAATAATTTTGGCCAAATTTTCGCATTTTTTGTAAGGGGTAACAT
CAT

The match between the T2 sequence and the C1/C2 sequence is

30

AAAAAAGTACCGCGTTTTACTCCTAATTACCAATTCTAACCATCCATATCA
CTTTTTGACGGACTCCGTGAAAATAATTTTTGGCCAAATTTTCGCATTTTTT
GTAAGGGGTAACATCATCAAAATTTGCGAAAAA

5

Example of an animal connectron – H. sapiens

10

All of the human genome that has been fully sequenced by both the NIH-lead global sequencing project and the Celera Genomics, Inc. project. The gene descriptors for this chromosome do not yet exist. Without the positions and directions of the genes, it is not possible to select from among the possible connectrons to determine the real connectrons.

15

Human chromosome 22 has been processed and there 31,000 possible connectrons.

The gene descriptors for all the chromosomes of the human genome should become available within the year.

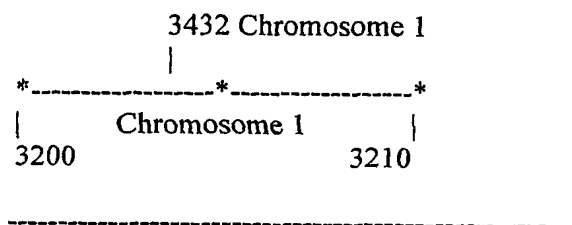
20

6. Permanent connectrons exist in prokaryotes, archæa, single-celled eukaryotes and multi-celled eukaryotes.

C1/C2 short loops are normally expressed as the 3'UTR of some gene. A class of connectron relationships exist that permit one C1/C2 short loop to control the existence of one or more T1-T2 long loops without being subject to any expression controls other than those of the gene to which the C1/C2 is 3'UTR. These connectron relationships are described as "permanent". Permanent connectrons exist in prokaryotes, archæa, single-celled eukaryotes and multi-celled eukaryotes.

Example of a prokaryote permanent connectron – E. coli

In this example the existence of the T1-T2 (3200-3210) long loop is controlled by a C1/C2 short loop (3432). The expression of this C1/C2 short loop is controlled only by the gene *btuB*.



A double stranded DNA loop of length 93.339 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 3200. This T1 control element has the DNA sequence

AAGCGGCACTGCTCTTTAACAATTTATCAGACAATCTGTGTGGGCACTCG
 AAGATACGGATTCTTAACGTCGCAAGACGAAAAATGAATACCAAGTCTCA
 AGAGTGAACACGTAATTCATTACGAAGTTTAATTCTTTGAGCATCAAACCTT
 TTAAATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCC
 TAACACATGCAAGTCGAACGGTAACAGGAAACAGCTTGCTGTTTCGCTGA
 CGAGTGGCGGACGGGTGAGTAATGTCTGGGAAACTGCCTGATGGAGGGG

GATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGCAAGACCAAA
GAGGGGGACCTTCGGGCCTCTTGCCATC

This double stranded DNA loop is bounded on the right by a T2 control element
whose identifier is 3310. This T2 control element has the DNA sequence

CAGACAATCTGTGTGGGCACTCGAAGATACGGATTCTTAACGTCGCAAGA
CGAAAAATGAATACCAAGTCTCAAGAGTGAACACGTAATTCATTACGAAG
TTTAATTCTTTGAGCGTCAAACCTTTTAAATTGAAGAGTTTGATCATGGCTC
AGATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACGGTAACA
GGAAGAAGCTTGCTTCTTTGCTGACGAGTGGCGGACGGGTGAGTAATGTC
TGGGAAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAAT
ACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCA
TCGGATGTGCCCAGATGGGATTAGCTAGT

This long T1/T2 double stranded DNA loop modulates the expression of the
following genes

rrsC	gltU	rrlC	rrfC	aspT	trpT	yifA	yifE	yifB
ilvL	ilvG_1	ilvM	ilvE	ilvD	ilvA	ilvY	ilvC	ppiC
b3776	rep	gppA	rhlB	trxA	rhoL	rho	rfe	wzzE
wecB	rffH	wecD	wecE	wzxE	yifM_2	wecG	yifK	
argX	hisR	leuT	proM	aslB	aslA	hemY	hemX	
hemD	cyaA	cyaY	b3808	dapF	uvrD	b3814	corA	
yigF	yigG	rarD	yigI	pldA	recQ	yigJ	yigK	pldB
yigL	yigM	metR	metE	ysgA	udp	yigN	ubiE	yigP
b3836	yigU	yigW_1	rfaH	yigC	ubiB	fadA	fadB	
pepQ	trkH	hemG						

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2
short loops.

A C1/C2 short loop on chromosome 1 whose identifier is 3432 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene *btuB* and has the DNA sequence

5 TGC GCGGTCAGAAAATTATTTTAAATTTCTCTTGTCAGGCCGGAATAACT
CCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGG
GTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATG
CTTGACTCTGTAGCGGGAAGGCGTATTATGCACACC...TGCAACTCGACTC
CATGAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATA
10 CGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGC
AAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCGCTTACCACTTTGTGAT
TCATGACTGGGGTGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGT
TGGATCACCTCCTTACCTTAAAGAAGCGT

15 The match between the T1 sequence and the C1/C2 sequence is

AAGCGGCACTGCTCTTTAACAATTTATCAGACAATCTGTGTGGGCACTCG
AAGATACGGATTCTTAACGTCGCAAGACGAAAAATGAATACCAAGTCTCA
AGAGTGAACACGTAATTCATTACGAAGTTAATTCTTTGAGC

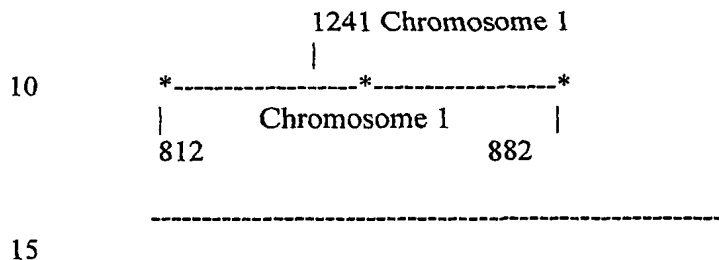
20

The match between the T2 sequence and the C1/C2 sequence is

CAGACAATCTGTGTGGGCACTCGAAGATACGGATTCTTAACGTCGCAAGA
CGAAAAATGAATACCAAGTCTCAAGAGTGAACACGTAATTCATTACGAAG
25 TTTAATTCTTTGAGCGTCAAACCTTTTAAATTGAAGAGTTTGATCATGGCTC
AGATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACGGTAACA
GGAAGAAGCTTGCTTCTTTGCTGACGAGTGGCGGACGGGTGAGTAATGTC
TGGGAAACTGCCTGATGGAGGGGGATACTACTGGAAACGGTAGCTAAT
ACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCA
30 TCGGATGTGCCCAGATGGGATTAGCTAGT

Example of an archea permanent connectron – *H. pylori*

In this example the existence of the T1-T2 (812-882) long loop is controlled by a C1/C2 short loop (1241). The expression of this C1/C2 short loop is controlled only by the gene HP1535.



A double stranded DNA loop of length 96.385 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 812. This T1 control element has the DNA sequence

20 TTTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 882. This T2 control element has the DNA sequence

25 TAGCGGAACTAAAGCATTCATCCCAAACACTAAAGATATTTGG

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

30

HP0999	HP1000	HP1001	HP1002	HP1003	HP1005	HP1006
HP1008	HP1009	HPtRNA-Pro	HP1010	HP1011	HP1013	HP1015
HP1017	HP1018	HP1020	HP1021	HP1022	HP1023	HP1024
HP1025	HP1027	HP1028	HP1030	HP1031	HP1033	HP1034
HP1038	HP1039	HP1040	HP1041	HP1042	HP1043	HP1044

HP1045	HP1046	HP1051	HP1052	HP1055	HP1056	HP1058
HP1060	HP1065	HPtRNA-Ser	HP1066	HP1067	HP1069	HP1070
HP1074	HP1075	HP1076	HP1077	HP1078	HP1079	HP1080
HP1081	HP1083	HP1084	HP1085	HP1088	HP1091	HP1092
5 HP1093	HP1094	HP1095	HP1096			

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

10 A C1/C2 short loop on chromosome 1 whose identifier is 1241 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene HP1535 and has the DNA sequence

15 TTTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCATTTCATCCC
AAACA

The match between the T1 sequence and the C1/C2 sequence is

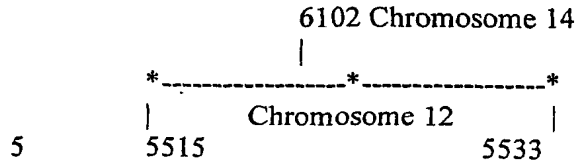
20 TTTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCA

The match between the T2 sequence and the C1/C2 sequence is

25 TAGCGGAACTAAAGCATTTCATCCCAAACA

Example of a single-celled permanent connectron – *S. cerevisiae*

30 In this example the existence of the T1-T2 (5515-5533) long loop is controlled by a C1/C2 short loop (6102). The expression of this C1/C2 short loop is controlled only by the gene YNL339C.



10 A double stranded DNA loop of length 6.466 kilo-bases on chromosome 12 is bounded on the left by a T1 sequence whose identifier is 5515. This T1 control element has the DNA sequence

15 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAC TGT TTTGATTTAGTGTTTGTTCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTG

20 This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 5533. This T2 control element has the DNA sequence

25 ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGA
ATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTTC
ACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATC
TATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACA
TCCGGGTAAGAGACAACAGGGCT

30 This long T1/T2 double stranded DNA loop modulates the expression of the following genes

YLR467W

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 14 whose identifier is 6102 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YNL339C and has the DNA sequence

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTGAATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

The match between the T1 sequence and the C1/C2 sequence is

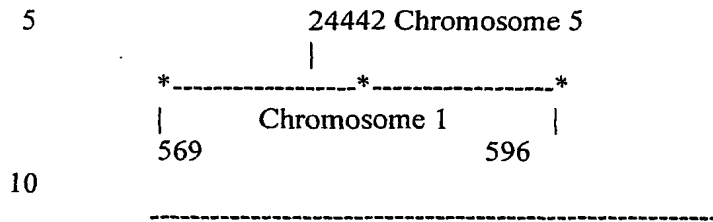
AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTAT
ATTGTAAGAAATTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTAT
TTCAGTGTGTTTGAATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTG
GGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAA
AGGTAGTAAGTAGCTTTTGGTTG

The match between the T2 sequence and the C1/C2 sequence is

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTCTAGGGA
ATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTC
ACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATC
TATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACA
TCCGGGTAAGAGACAACAGGGCT

Example of a multi-celled permanent connectron – *C. elegans*

In this example the existence of the T1-T2 (5515-5533) long loop is controlled by a C1/C2 short loop (6102). The expression of this C1/C2 short loop is controlled only by the gene YNL339C.



A double stranded DNA loop of length 30.606 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 569. This T1 control element has the DNA sequence

AAATCGAGCCCGTAAATCGACACAAGCGCTACAGTAGTC

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 596. This T2 control element has the DNA sequence

AGTGCTACAGTAGTCATTAAAGAATTACTGTAGTTTTTCGCT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 5 whose identifier is 24442 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene F20D6.4 and has the DNA sequence

30 GAGCCCGTAAATCGACACAAGCGCTACAGTAGTCATTAAAGAATTACTG
TAGTTTTTC

The match between the T1 sequence and the C1/C2 sequence is

GAGCCCGTAAATCGACACAAGCGCTACAGTAGTC

The match between the T2 sequence and the C1/C2 sequence is

5

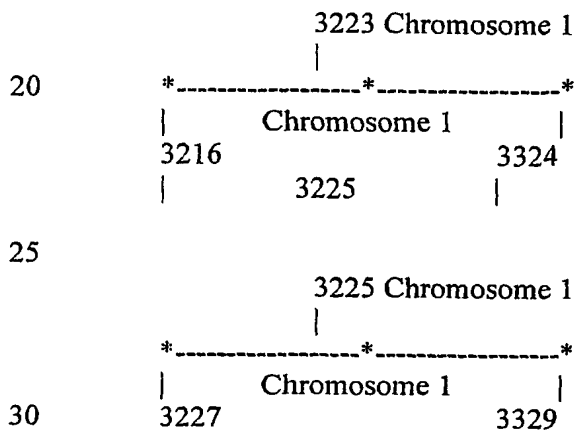
GCTACAGTAGTCATTTAAAGAATTACTGTAGTTTTC

7. Transient connectrons exist in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes.

5 A class of connectron relationships exist that permit one C1/C2 short loop to control the existence of one or more T1-T2 long loops such that this C1/C2 short loop is itself subject to expression control by another T1-T2 long loop which surrounds it. These connectron relationships are described as "transient". Transient connectrons exist in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes.

10 Example of a prokaryote transient connectron – E. coli

In this example the existence of the T1-T2 (3227-3329) long loop is controlled by the C1/C2 (3225) short loop. The expression of this C1/C2 short loop is controlled by the existence of the T1-T2 (3216-3324) long loop. The existence of this T1-T2 long loop is itself determined by the expression of the C1/C2 (3223) short loop. The C1/C2 (3225) short loop is the transient connectron.



35 A double stranded DNA loop of length 93.464 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 3216. This T1 control element has the DNA sequence

AGCGCAAGCGAAGCTCTTGATCGAAGCCCCGGTAAACGGCGGCCGTA
 ATAACGGTCCTAAGGTAGCGAAATTCCTTGTCGGGTAAGTTCCGACCTGC
 ACGAATGGCGTAATGATGGCCAGGCTGTCTCCACCCGAGACTCAGTGAAA
 TTGAACTCGCTGTGAAGATGCAGTGTACCCGCGGCAAGACGGAAAGACCC
 5 CGTGAACCTTTACTATAGCTTGACACTGAACATTGAGCCTTGATGTGTAGG
 ATAGGTGGGAGGCTTTGAAGTGTGGACGCCAGTCTGCATGGAGCCGACCT
 TGAAATACCACCCTTTAATGTTTGATGTTCTAACGT

10 This double stranded DNA loop is bounded on the right by a T2 control element
 whose identifier is 3324. This T2 control element has the DNA sequence

CCCGGTAAACGGCGGCCGTAACCTATAACGGTCCTAAGGTAGCGAAATTCC
 TTGTCGGGTAAGTTCCGACCTGCACGAATGGCGTAATGATGGCCAGGCTG
 TCTCCACCCGAGACTCAGTGAAATTGAACTCGCTGTGAAGATGCAGTGTA
 15 CCCGCGGCAAGACGGAAAGACCCCGTGAACCTTTACTATAGCTTGACACT
 GAACATTGAGCCTTGATGTGTAGGATAGGTGGGAGGCTTTGAAGTGTGGA
 CGCCAGTCTGCATGGAGCCGACCTTGAAATACCACCCTTTAATGTTTGATG
 TTCTAACGTTGACCCGTAATCCGGGTTGCGGACAGT

20 This long T1/T2 double stranded DNA loop modulates the expression of the
 following genes

	rrfC	aspT	trpT	yifA	yifE	yifB	ilvL	ilvG_1	ilvM
	ilvE	ilvD	ilvA	ilvY	ilvC	ppiC	b3776	rep	gppA
25	rhlB	trxA	rhoL	rho	rfe	wzzE	wecB	rffH	wecD
	wecE	wzxE	yifM_2	wecG	yifK	argX	hisR	leuT	
	proM	aslB	aslA	hemY	hemX	hemD	cyaA	cyaY	
	b3808	dapF	uvrD	b3814	corA	yigF	yigG	rarD	yigI
	pIdA	recQ	yigJ	yigK	pIdB	yigL	yigM	metR	metE
30	ysgA	udp	yigN	ubiE	yigP	b3836	yigU	yigW_1	rfaH
	yigC	ubiB	fadA	fadB	pepQ	trkH	hemG	rrsA	ileT
	rrlA								

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

5 A C1/C2 short loop on chromosome 1 whose identifier is 3225 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene *rrlC* and has the DNA sequence

10 AAACAGAATTTGCCTGGCGGCCGTAGCGCGGTGGTCCCACCTGACCCCAT
GCCGAAGTCTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTC
CCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATTA

15 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is 3323 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene *rrlA* and has the DNA sequence

20 GCGAAGCTCTTGATCGAAGCCCCGGTAAACGGCGGCCGTAACTATAACGG
TCCTAAGGTAGCGAAATTCCTTGTCGGGTAAAGTTCCGACCTGCACGAATG
GCGTAATGATGGCCAGGCTGTCTCCACCCGAGACTCAGTGAAATTGAACT
CGCTGTGAAGATGCAGTGTACCCGCGGCAAGACGGA...AACAGAATTTGC
25 CTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAAGTCTAGAA
GTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTC

The match between the T1 sequence and the C1/C2 sequence is

30 GCGAAGCTCTTGATCGAAGCCCCGGTAAACGGCGGCCGTAACTATAACGG
TCCTAAGGTAGCGAAATTCCTTGTCGGGTAAAGTTCCGACCTGCACGAATG
GCGTAATGATGGCCAGGCTGTCTCCACCCGAGACTCAGTGAAATTGAACT

CGCTGTGAAGATGCAGTGTACCCGCGGCAAGACGGAAAGACCCCGTGAA
 CCTTTACTATAGCTTGACACTGAACATTGAGCCTTGATGTGTAGGATAGGT
 GGGAGGCTTTGAAGTGTGGACGCCAGTCTGCATGGAGCCGACCTTGAAAT
 ACCACCCTTTAATGTTTGATGTTCTAACGT

5

The match between the T2 sequence and the C1/C2 sequence is

CCCGGTAAACGGCGGCCGTAACCTATAACGGTCCTAAGGTAGCGAAATTCC
 TTGTTCGGGTAAGTTCCGACCTGCACGAATGGCGTAATGATGGCCAGGCTG
 TCTCCACCCGAGACTCAGTGAAATTGAACTCGCTGTGAAGATGCAGTGTA
 CCCGCGGCAAGACGGAAAGACCCCGTGAACCTTTACTATAGCTTGACACT
 GAACATTGAGCCTTGATGTGTAGGATAGGTGGGAGGCTTTGAAGTGTGGA
 CGCCAGTCTGCATGGAGCCGACCTTGAAATACCACCCTTTAATGTTTGATG
 TTCTAACGTTGACCCGTAATCCGGGTTGCGGACAGT

10

15

A double stranded DNA loop of length 93.749 kilo-bases on chromosome 1 is
 bounded on the left by a T1 sequence whose identifier is 3227. This T1 control
 element has the DNA sequence

20

AGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCA
 GG

25

This double stranded DNA loop is bounded on the right by a T2 control element
 whose identifier is 3329. This T2 control element has the DNA sequence

CATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAG
 TCG

30

This long T1/T2 double stranded DNA loop modulates the expression of the
 following genes

	aspT	trpT	yifA	yifE	yifB	ilvL	ilvG_1	ilvM	ilvE
	ilvD	ilvA	ilvY	ilvC	ppiC	b3776	rep	gppA	rhlB
	trxA	rhoL	rho	rfe	wzzE	wecB	rffH	wecD	wecE
5	wzxE	yifM_2	wecG	yifK	argX	hisR	leuT	proM	
	aslB	aslA	hemY	hemX	hemD	cyaA	cyaY	b3808	
	dapF	uvrD	b3814	corA	yigF	yigG	rarD	yigI	pldA
	recQ	yigJ	yigK	pldB	yigL	yigM	metR	metE	ysgA
	udp	yigN	ubiE	yigP	b3836	yigU	yigW_1	rfaH	yigC
10	ubiB	fadA	fadB	pepQ	trkH	hemG	rrsA	ileT	rrlA
	rrfA								

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

15

A C1/C2 short loop on chromosome 1 whose identifier is 3225 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene rrlC and has the DNA sequence

20

AAACAGAATTTGCCTGGCGGCCGTAGCGCGGTGGTCCCACCTGACCCCAT
GCCGAACCTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTC
CCCATGCGAGAGTAGGGAAGTCCAGGCATCAAATTA

The match between the T1 sequence and the C1/C2 sequence is

25

AGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAAGTGCCA
GG

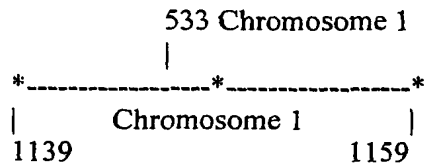
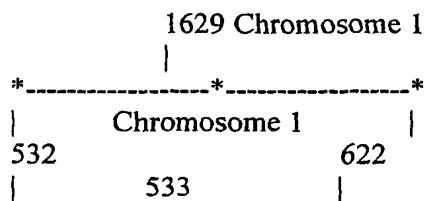
The match between the T2 sequence and the C1/C2 sequence is

30

CATGCGAGAGTAGGGAAGTCCAGGCATCAAAT

Example of an archea transient connectron – *M. jannaschii*

5 In this example the existence of the T1-T2 (1139-1159) long loop is controlled by the C1/C2 (533) short loop. The expression of this C1/C2 short loop is controlled by the existence of the T1-T2 (532-622) long loop. The existence of this T1-T2 long loop is itself determined by the expression of the C1/C2 (1629) short loop. The C1/C2 (533) short loop is the transient connectron.



25 -----

A double stranded DNA loop of length 78.672 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 532. This T1 control element has the DNA sequence

30 ATATGTTTGAAATTTGAAAATAAGAGTATTTAG

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 622. This T2 control element has the DNA sequence

35 TTGAAAATAAGAGCATTTAGAAGTTATTAATTAGTTCAAAGGATTTT

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

	MJ0486	MJ0487	MJ0488	MJ0489	MJ0490	MJ0492	MJ0493
5	MJ0494	MJ0495	MJ0496	MJ0497	MJ0499	MJ0500	MJ0501
	MJ0502	MJ0503	MJ0504	MJ0506	MJ0507	MJ0508	MJ0509
	MJ0510	MJ0511	MJ0512	MJ0513	MJ0514	MJ0514	MJ0517
	MJ0519	MJ0520	MJ0521	MJ0522	MJ0523	MJ0525	MJ0526
	MJ0526	MJ0529	MJ0530	MJ0531	MJ0532	MJ0534	MJ0535
10	MJ0536	MJ0538	MJ0539	MJ0540	MJ0541	MJ0542	MJ0543
	MJ0544	MJ0545	MJ0547	MJ0548	MJ0549	MJ0550	MJ0552
	MJ0553	MJ0554	MJ0555	MJ0556	MJ0557	MJ0558	MJ0559
	MJ0560	MJ0561	MJ0562	MJ0563	MJ0564	MJ0565	MJ0566
	MJ0568	MJ0569	MJ0570				

15

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

20 A C1/C2 short loop on chromosome 1 whose identifier is 533 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ0485 and has the DNA sequence

25 ATTTTATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGA
GTTTATTGAATT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

30 A C1/C2 short loop on chromosome 1 whose identifier is 1629 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1597 and has the DNA sequence

ATATGTTTGAAATTTGAAAATAAGAGTATTTAGAAGTTATTAATTAGTTCA
AAGGATTTTTATTTAATTTCTAAGGGTTTGCTGGTTTGATTATTTAGAATAT
TTGAGTTTATTGAATTATTCAGATTTTAAAAAATTA

5

The match between the T1 sequence and the C1/C2 sequence is

ATATGTTTGAAATTTGAAAATAAGAGTATTTAG

10

The match between the T2 sequence and the C1/C2 sequence is

ATTTAGAAGTTATTAATTAGTTCAAAGGATTTT

15

A double stranded DNA loop of length 14.509 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 1139. This T1 control element has the DNA sequence

20

ATTTATTAATTAGTTCAAAGGATTTTTATTTAATTTCTAAGGGTTAGCTGG
TTTGATTGTTTAAAATATTTGAGTTTA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 1159. This T2 control element has the DNA sequence

25

ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTAT
TGAATTATTCAGATTTTAAAAAATTA

30

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

MJ1096 MJ1097 tRNA-Arg-3 MJ1098 MJ1099 MJ1100 MJ1101
 MJ1102 MJ1103 MJ1104 MJ1105 MJ1106 MJ1107 MJ1108

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is 533 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ0485 and has the DNA sequence

ATTTTATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGA
 GTTTATTGAATT

The match between the T1 sequence and the C1/C2 sequence is

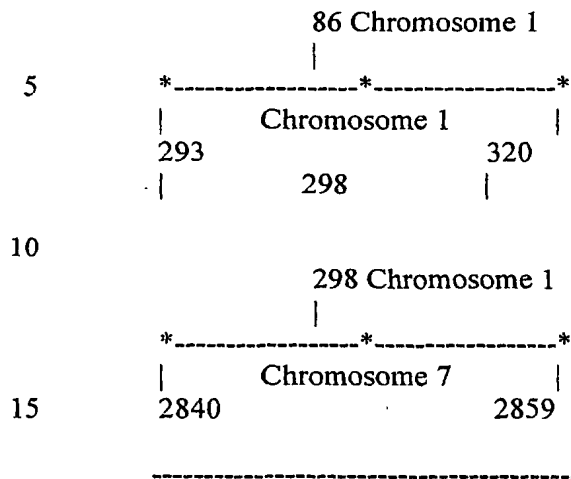
ATTTTATTTAATTTCTAAGGGTTAGCTGGTTTGATT

The match between the T2 sequence and the C1/C2 sequence is

ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTAT
 TGAATT

Example of a single-celled transient connectron – *S. cerevisiae*

In this example the existence of the T1-T2 (2840-2859) long loop is controlled by the C1/C2 (298) short loop. The expression of this C1/C2 short loop is controlled by the existence of the T1-T2 (293-320) long loop. The existence of this T1-T2 long loop is itself determined by the expression of the C1/C2 (86) short loop. The C1/C2 (298) short loop is the transient connectron.



A double stranded DNA loop of length 38.470 kilo-bases on chromosome 2 is bounded on the left by a T1 sequence whose identifier is 293. This T1 control element has the DNA sequence

GAATTGTTGGAATAAAAATCCACTATCGTCTATCAACTAATAGTTATATTA
TCAATATATTATCATATACGGTGTTAAGATGATGACATAAGTTATGAGAA
GCTGTCATCGAAGTTAGAGGAAGCTGAAGTGCAAGGATTGATAATGTAAT
AGGATAATGAAACATATAAAACGGAATGAGGAATAATCGTAATATTAGT
ATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTTGAGGAGAAC
TTCTAGT

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 320. This T2 control element has the DNA sequence

AATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTC
GAGGAGAACTTCTAGTATATTCTGTA

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

YBL005W-B TS(AGA)B YBL004W YBL003C YBL002W YBL001C
YBR001C YBR002C YBR003W YBR004C YBR005W YBR006W
YBR007C YBR008C YBR009C YBR010W YBR011C YBR012C

5

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

10 A C1/C2 short loop on chromosome 2 whose identifier is 298 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YBL005W-B and has the DNA sequence

15 ATCTATTACATTATGGGTGGTATGTTGGAATAAAAATCCACTATCGTCTAT
CAACTAATAGTTATATTATCAATATATTATCATATACGGTGTTAAGATGAT
GACATAAGTTATGAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAAGTGCA
AGGATTGATAATGTAATAGGATAATGAAACATATAAAACGGAATGAGGA
ATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCT
ATATCCTTGAGGAGAACTTCTAGTATATTCTGTATACCTAATATTATAGCC
20 TTTATCAACAATGGAATCCCAACAATTATCTCAACATTC

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

25 A C1/C2 short loop on chromosome 1 whose identifier is 86 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YAR009C and has the DNA sequence

30 ATCTATTACATTATGGGTGGTATGTTGGAATAGAAATCAACTATCATCTAC
TAACTAGTATTTACATTACTAGTATATTATCATATACGGTGTTAGAAGATG
ACGCAAATGATGAGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGCA
AGGATTGATAATGTAATAGGATCAATGAATATAAACATATAAAACGGAAT

GAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGG
ATTCCTATATCCTCGAGGAGAACTTCTAGTATATTCTGTATACCTAATATT
ATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAACATTCACCCAT
TTCTCAGAA

5

The match between the T1 sequence and the C1/C2 sequence is

AAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAAT
ATAGATTCCATTTTGAGGATTCCTATATCCT

10

The match between the T2 sequence and the C1/C2 sequence is

AATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTC
GAGGAGAACTTCTAGTATATTCTGTA

15

A double stranded DNA loop of length 5.302 kilo-bases on chromosome 7 is bounded
on the left by a T1 sequence whose identifier is 2840. This T1 control element has
the DNA sequence

20

TCTGTTGGAATAAAAATCCACTATCGTCTATCAACTAATAGTTATATTATC
AATATATTATCATATACGGTGTTAAGATGATGACATAAGTTATGAGAAGC
TGTCATCGAAGTTAGAGGAAGCTGAAACGCAAGGATTGATAATGTAATAG
GATCAATGAATATAAACATATAAAACGGAATGAGGAATAATCGTAATATT
AGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAG
AACTTCTAGTATATTCTGTATACCTAAATTATAGCCTTTATCAACAATGGA
ATCCCAACAA

25

30

This double stranded DNA loop is bounded on the right by a T2 control element
whose identifier is 2859. This T2 control element has the DNA sequence

CTATCAACTAATAGTTATATTATCAATATATTATCATATACGGTGTTAAGA
TGATGACATAAGTTATGAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAAA
CGCAAGGATTGATAATGTAATAGGATCAATGAATATAAACATATAAAACG
GAATGAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTT
5 GAGGATTCCTATATCCTCGAGGAGAACTTCTAGTATATTCTGTATACCTAA
TATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAACATTCAC
ATATTTCTCAT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2
10 short loops.

A C1/C2 short loop on chromosome 2 whose identifier is 298 controls the expression
of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene YBL005W-B and has the DNA sequence

15 ATCTATTACATTATGGGTGGTATGTTGGAATAAAAATCCACTATCGTCTAT
CAACTAATAGTTATATTATCAATATATTATCATATACGGTGTTAAGATGAT
GACATAAGTTATGAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAAGTGCA
AGGATTGATAATGTAATAGGATAATGAAACATATAAAACGGAATGAGGA
20 ATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCT
ATATCCTTGAGGAGAACTTCTAGTATATTCTGTATACCTAATATTATAGCC
TTTATCAACAATGGAATCCCAACAATTATCTCAACATTC

The match between the T1 sequence and the C1/C2 sequence is

25 TGTTGGAATAAAAATCCACTATCGTCTATCAACTAATAGTTATATTATCAA
TATATTATCATATACGGTGTTAAGATGATGACATAAGTTATGAGAAGCTG
TCATCGAAGTTAGAGGAAGCTGAA

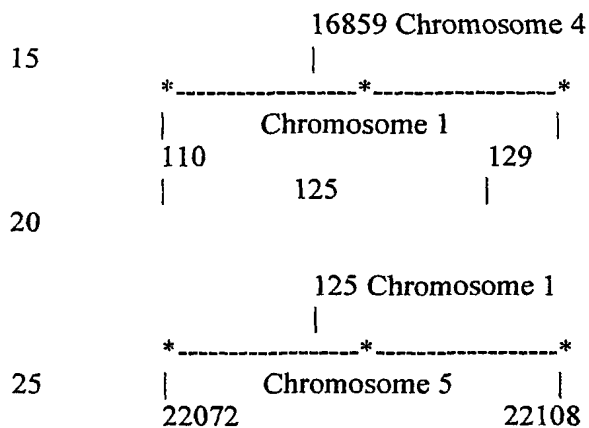
30 The match between the T2 sequence and the C1/C2 sequence is

CTATCAACTAATAGTTATATTATCAATATATTATCATATACGGTGTTAAGA
 TGATGACATAAGTTATGAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAA

5

Example of a multi-celled transient connectron – *C. elegans*

In this example the existence of the T1-T2 (22072-22108) long loop is controlled by the C1/C2 (125) short loop. The expression of this C1/C2 short loop is controlled by the existence of the T1-T2 (110-129) long loop. The existence of this T1-T2 long loop is itself determined by the expression of the C1/C2 (16859) short loop. The C1/C2 (125) short loop is the transient connectron.



A double stranded DNA loop of length 18.855 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 110. This T1 control element has the DNA sequence

AGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGGC

35

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 129. This T2 control element has the DNA sequence

TTCTCCCGCATTTTTTGTAGATCTACGTAGATCAAACCGAAATGAGGCACT
 TTCTGAATCCACGAGCTAGGCTTAAGCTTAGGCTTAAGCTTAGGCCTTTTC
 TCAGGCTTAGGCTTAGGCTTA

5

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

ZC123.3 ZC123.2

10

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

15

A C1/C2 short loop on chromosome 1 whose identifier is 125 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ZC123.3 and has the DNA sequence

20

ACGCGCCGTAAATCTACCCCAGATATGGCCGAGCCAAAATGGCCTAGTTC
 GGCAAACCTCTTTCATTTCAATTTATGAGGGAAGCCAGAA

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

25

A C1/C2 short loop on chromosome 4 whose identifier is 16859 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene F58E2.7 and has the DNA sequence

30

CTTAGGCTTAAGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGGCTTAAGCTT
 AGGCTTAAGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGGCTTAAGCTTAG
 GCTTAAGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGGCT
 TAAGCTTAGACTTA

The match between the T1 sequence and the C1/C2 sequence is

AGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGGC

5

The match between the T2 sequence and the C1/C2 sequence is

TAGGCTTAAGCTTAGGCTTAAGCTTAGGC

10

A double stranded DNA loop of length 51.031 kilo-bases on chromosome 5 is bounded on the left by a T1 sequence whose identifier is 22072. This T1 control element has the DNA sequence

15

CGCAACGCGCCGTAAATCTACCCCAGATATGGCCGAGCCAAAATGACCTA
GTTCGGC

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 22108. This T2 control element has the DNA sequence

20

TGACAATCGCCTGCCGGACAACGCGTGGAAGTGTCGTGTACTCCACAC
GGACAAATACATTTAGTTTTACAATAAAATCGAACCGCGACGCGACACG
CAACGCGACGTAAATCTACCCCAGATATGGCCGAGCCAAAATGGCCTAGT
TCGGCAAACCTCTTCTATTTC

25

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

30

F36H9.3 F36H9.4 F36H9.5 F36H9.2 F36H9.1 F36H9.6

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

5 A C1/C2 short loop on chromosome 1 whose identifier is 125 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ZC123.3 and has the DNA sequence

ACGCGCCGTAAATCTACCCCAGATATGGCCGAGCCAAAATGGCCTAGTTC
10 GGCAAACCTCTTTCATTTCAATTTATGAGGGAAGCCAGAA

The match between the T1 sequence and the C1/C2 sequence is

ACGCGCCGTAAATCTACCCCAGATATGGCCGAGCCAAAATG

15 The match between the T2 sequence and the C1/C2 sequence is

CGTAAATCTACCCCAGATATGGCCGAGCCAAAATGGCCTAGTTCGGCAAA
CTCTT

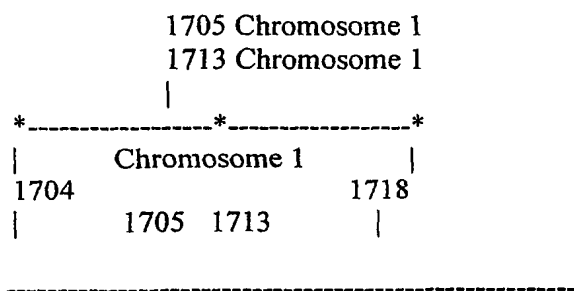
20 -----

8. Self-limiting connectrons occur in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes

A class of connectron relationships exist that permit one C1/C2 short loop to control the existence of the T1-T2 long loop that surrounds it. These connectron relationships are described as "self-limiting". Self-limiting connectrons exist in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes.

Example of a prokaryotic self-limiting connectrons – E. coli

In this example the existence of the T1-T2 (1704-1718) long loop is controlled by two C1/C2 (1705 and 1713) short loops. The expression of these C1/C2 short loops is controlled by the existence of the T1-T2 (1704-1718) long loop. The existence of this T1-T2 long loop is itself determined by the expression of the two C1/C2 (1705 and 1713) short loops. The C1/C2 (1705 and 1713) short loops are the self-limiting connectrons.



A double stranded DNA loop of length 15.259 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 1704. This T1 control element has the DNA sequence

CGCCCCGTTACACGATTCTCTGTAGTTCAGTCGGTAGAACGGCGGACT
GTTAATCCGTATGTCACTGGT

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 1718. This T2 control element has the DNA sequence

5 TTCAGTCGGTAGAACGGCGGACTGTTAATCCGTATGTCACTGGTTCGAGTC
CAGTCAGAGGAGCCAAATTC

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

10 asnT b1978 b1979 b1980 shiA amn b1983 asnW
yeeO asnU

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

15

A C1/C2 short loop on chromosome 1 whose identifier is 1705 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene and has the DNA sequence

20

CGCCCCGTTACACGATTCCCTCTGTAGTTCAGTCGGTAGAACGGCGGACT
GTTAATCCGTATGTCACTGGTTCGAGTCCAGTCAGAGGAGCCAAATTC

25

A C1/C2 short loop on chromosome 1 whose identifier is 1713 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene asnW and has the DNA sequence

30

CACGATTCCCTCTGTAGTTCAGTCGGTAGAACGGCGGACTGTTAATCCGTAT
GTCACTGGTTCGAGTCCAGTCAGAGGAGCCAAATT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

5 A C1/C2 short loop on chromosome 1 whose identifier is 1705 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene and has the DNA sequence

CGCCCCGTTACACGATTCCTCTGTAGTTCAGTCGGTAGAACGGCGGACT
GTTAATCCGTATGTCACTGGTTCGAGTCCAGTCAGAGGAGCCAAATTC

10 The match between the T1 sequence and the C1/C2 sequence is

CGCCCCGTTACACGATTCCTCTGTAGTTCAGTCGGTAGAACGGCGGACT
GTTAATCCGTATGTCACTGGT

15 The match between the T2 sequence and the C1/C2 sequence is

TTCAGTCGGTAGAACGGCGGACTGTTAATCCGTATGTCACTGGTTCGAGTC
CAGTCAGAGGAGCCAAATTC

20 A C1/C2 short loop on chromosome 1 whose identifier is 1713 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene asnW and has the DNA sequence

25 CACGATTCCTCTGTAGTTCAGTCGGTAGAACGGCGGACTGTTAATCCGTAT
GTCACTGGTTCGAGTCCAGTCAGAGGAGCCAAATT

The match between the T1 sequence and the C1/C2 sequence is

30 CACGATTCCTCTGTAGTTCAGTCGGTAGAACGGCGGACTGTTAATCCGTAT
GTCACTGGT

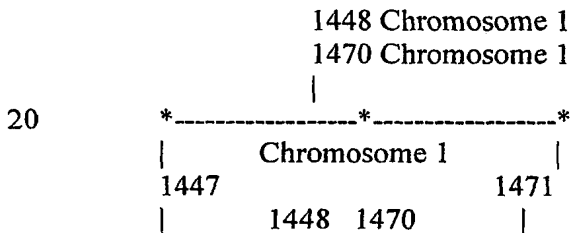
The match between the T2 sequence and the C1/C2 sequence is

TTCAGTCGGTAGAACGGCGGACTGTTAATCCGTATGTCACTGGTTCGAGTC
CAGTCAGAGGAGCCAAATT

5

Example of a archea self-limiting connectrons – *M. jannaschii*

10 In this example the existence of the T1-T2 (1447-1471) long loop is controlled by
two C1/C2 (1448 and 1470) short loops. The expression of these C1/C2 short loops is
controlled by the existence of the T1-T2 (1447-1471) long loop. The existence of this
T1-T2 long loop is itself determined by the expression of the two C1/C2 (1705 and
1713) short loops. The C1/C2 (1448 and 1470) short loops are the self-limiting
15 connectrons.



A double stranded DNA loop of length 22.675 kilo-bases on chromosome 1 is
bounded on the left by a T1 sequence whose identifier is 1447. This T1 control
element has the DNA sequence

30

TTATAGAACATTATGAAGCTTTTTACTCAACTAACAACCGTATCGAATTTA
CCATTACTTGGAAATCTATTTAAAACCTCTTTAATCTTATGATA

35

This double stranded DNA loop is bounded on the right by a T2 control element
whose identifier is 1471. This T2 control element has the DNA sequence

CAACTAACAACCGTATCGAATTTACCATTACTTGGAAATCTATTTAAAACC
TCTTTAATCTTGTGATAATAAATTCTAATCGATTTCGTGACTTAT

5 This long T1/T2 double stranded DNA loop modulates the expression of the following genes

	MJ1402	MJ1403	MJ1404	MJ1405	MJ1406	MJ1407	MJ1408
	MJ1409	MJ1410	MJ1411	MJ1412	MJ1413	MJ1414	MJ1415
10	MJ1416	MJ1417	MJ1418	MJ1419	MJ1420		

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

15 A C1/C2 short loop on chromosome 1 whose identifier is 1448 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1401 and has the DNA sequence

20 TTATAGAACATTATGAAGCTTTTTACTCAACTAACAACCGTATCGAATT
CCATTACTTGGAAATCTATTTAAAACCTCTTTAATCTTATGATAATAAATT
CTAATCGATTTCGTGACTTAT

25 A C1/C2 short loop on chromosome 1 whose identifier is 1470 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1420 and has the DNA sequence

30 TTATAGAACATTATGAAGCTTTTTACTCAACTAACAACCGTATCGAATT
CCATTACTTGGAAATCTATTTAAAACCTCTTTAATCTTGTGATAATAAATT
CTAATCGATTTCGTG

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

5 A C1/C2 short loop on chromosome 1 whose identifier is 1470 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1420 and has the DNA sequence

10 TTATAGAACATTATGAAGCTTTTTACTCAACTAACAACCGTATCGAATTTA
CCATTACTTGGAAATCTATTTAAAACCTCTTTAATCTTGTGATAATAAATT
CTAATCGATTCGTG

The match between the T1 sequence and the C1/C2 sequence is

15 TTATAGAACATTATGAAGCTTTTTACTCAACTAACAACCGTATCGAATTTA
CCATTACTTGGAAATCTATTTAAAACCTCTTTAATCTT

The match between the T2 sequence and the C1/C2 sequence is

20 CAACTAACAACCGTATCGAATTTACCATTACTTGGAAATCTATTTAAAACC
TCTTTAATCTTGTGATAATAAATTCTAATCGATTCGTG

25 A C1/C2 short loop on chromosome 1 whose identifier is 1448 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1401 and has the DNA sequence

TTATAGAACATTATGAAGCTTTTTACTCAACTAACAACCGTATCGAATTTA
CCATTACTTGGAAATCTATTTAAAACCTCTTTAATCTTATGATAATAAATT
CTAATCGATTCGTGACTTAT

30 The match between the T1 sequence and the C1/C2 sequence is

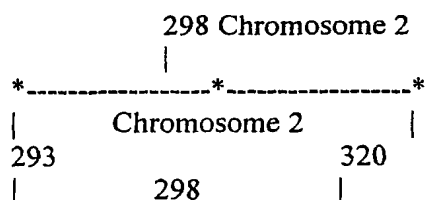
TTATAGAACATTATGAAGCTTTTTACTCAACTAACAACCGTATCGAATTTA
CCATTACTTGGAAATCTATTTAAAACCTCTTTAATCTTATGATA

The match between the T2 sequence and the C1/C2 sequence is

CAACTAACAACCGTATCGAATTTACCATTACTTGGAAATCTATTTAAAACC
TCTTTAATCTT

Example of a single-celled self-limiting connectron – *S. cerevisiae*

In this example the existence of the T1-T2 (293-320) long loop is controlled by C1/C2 (298) short loop. The expression of this C1/C2 short loop is controlled by the existence of the T1-T2 (293-320) long loop. The existence of this T1-T2 long loop is itself determined by the expression of the C1/C2 (298) short loop. The C1/C2 (298) short loop is the self-limiting connectron.



A double stranded DNA loop of length 38.470 kilo-bases on chromosome 2 is bounded on the left by a T1 sequence whose identifier is 293. This T1 control element has the DNA sequence

GAATTGTTGGAATAAAAATCCACTATCGTCTATCAACTAATAGTTATATTA
TCAATATATTATCATATACGGTGTTAAGATGATGACATAAGTTATGAGAA
GCTGTCATCGAAGTTAGAGGAAGCTGAAGTGCAAGGATTGATAATGTAAT

AGGATAATGAAACATATAAAACGGAATGAGGAATAATCGTAATATTAGT
ATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTTGAGGAGAAC
TTCTAGT

- 5 This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 320. This T2 control element has the DNA sequence

AATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTC
GAGGAGAACTTCTAGTATATTCTGTA

10

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

- YBL005W-B TS(AGA)B YBL004W YBL003C YBL002W YBL001C
15 YBR001C YBR002C YBR003W YBR004C YBR005W YBR006W
YBR007C YBR008C YBR009C YBR010W YBR011C YBR012C

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

20

A C1/C2 short loop on chromosome 2 whose identifier is 298 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YBL005W-B and has the DNA sequence

25

ATCTATTACATTATGGGTGGTATGTTGGAATAAAAAATCCACTATCGTCTAT
CAACTAATAGTTATATTATCAATATATTATCATATACGGTGTTAAGATGAT
GACATAAGTTATGAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAAGTGCA
AGGATTGATAATGTAATAGGATAATGAAACATATAAAACGGAATGAGGA
30 ATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCT
ATATCCTTGAGGAGAACTTCTAGTATATTCTGTATACCTAATATTATAGCC
TTTATCAACAATGGAATCCCAACAATTATCTCAACATTC

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

- 5 A C1/C2 short loop on chromosome 2 whose identifier is 298 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YBL005W-B and has the DNA sequence

10 ATCTATTACATTATGGGTGGTATGTTGGAATAAAAATCCACTATCGTCTAT
CAACTAATAGTTATATTATCAATATATTATCATATACGGTGTTAAGATGAT
GACATAAGTTATGAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAAGTGCA
AGGATTGATAATGTAATAGGATAATGAAACATATAAAACGGAATGAGGA
ATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCT
15 ATATCCTTGAGGAGAACTTCTAGTATATTCTGTATACCTAATATTATAGCC
TTTATCAACAATGGAATCCCAACAATTATCTCAACATTC

The match between the T1 sequence and the C1/C2 sequence is

20 TGTGGAATAAAAATCCACTATCGTCTATCAACTAATAGTTATATTATCAA
TATATTATCATATACGGTGTTAAGATGATGACATAAGTTATGAGAAGCTG
TCATCGAAGTTAGAGGAAGCTGAAGTGCAAGGATTGATAATGTAATAGGA
TAATGAAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTA
GAAATATAGATTCCATTTTGAGGATTCCTATATCCTTGAGGAGAACTTCTA
GT

25

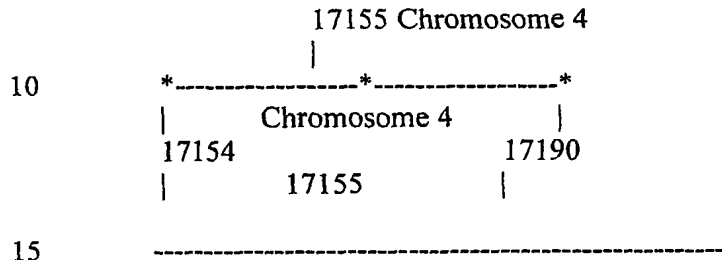
The match between the T2 sequence and the C1/C2 sequence is

AATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCT

30

Example of a multi-celled self-limiting connectron – *C. elegans*

In this example the existence of the T1-T2 (293-320) long loop is controlled by C1/C2 (298) short loop. The expression of this C1/C2 short loop is controlled by the existence of the T1-T2 (293-320) long loop. The existence of this T1-T2 long loop is itself determined by the expression of the C1/C2 (298) short loop. The C1/C2 (298) short loop is the self-limiting connectron.



A double stranded DNA loop of length 89.919 kilo-bases on chromosome 4 is bounded on the left by a T1 sequence whose identifier is 17154. This T1 control element has the DNA sequence

AAATTTCGGCAAATCGGCAAACCTGGCAA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 17190. This T2 control element has the DNA sequence

AATTGCGGATTTGCCGAATTTGTCTGACA

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

R08C7.11 M01H9.2 M01H9.3 M01H9.4 M01H9.1 ZK180.1 ZK180.2
ZK180.3 ZK180.4 ZK180.5 ZK180.6 ZK185.3 ZK185.2

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

5 A C1/C2 short loop on chromosome 4 whose identifier is 17155 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene R08C7.1 and has the DNA sequence

10 AAATTTCCGGCAAATCGGCAAACCTGGCAATTTGCCGATTTGCCGAATTTGT
CGACA

15 A C1/C2 short loop on chromosome 4 whose identifier is 17171 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ZK180.2 and has the DNA sequence

TGGAAATTTTCAGAATTTCAATTTTAATCGGCAAAATTGTACGCATCCTATG
AATTT

20 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

25 A C1/C2 short loop on chromosome 4 whose identifier is 17155 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene R08C7.1 and has the DNA sequence

AAATTTCCGGCAAATCGGCAAACCTGGCAATTTGCCGATTTGCCGAATTTGT
CGACA

30 The match between the T1 sequence and the C1/C2 sequence is

AAATTTCCGGCAAATCGGCAAACCTGGCAA

The match between the T2 sequence and the C1/C2 sequence is

AATTTGCCGATTTGCCGAATTTGTCGACA

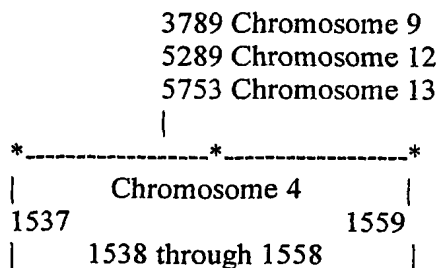
5

9. Geneless connectrons exist in single-celled and multi-celled eukaryotes

Normally T1-T2 long loops contain genes whose expression is regulated by the existence of the long loop. When a T1-T2 long loop does not contain any genes it is described as being "geneless". The existence of the T1-T2 long loop is itself controlled by one or more C1/C2 short loops that may be on the same or different chromosomes. The geneless T1-T2 long loops must contain one or more C1/C2 short loops.

Example of a single-celled geneless connectron – *S. cerevisiae*

In this example the existence of the T1-T2 (1537-1559) long loop is controlled by three C1/C2 (3789, 5289 and 5753) short loops. The expression of 21 C1/C2 (1538 through 1558) short loops are controlled by the existence of the T1-T2 (1537-1559) long loop.



A double stranded DNA loop of length 4.825 kilo-bases on chromosome 4 is bounded on the left by a T1 sequence whose identifier is 1537. This T1 control element has the DNA sequence

ATGAGATATATGTGGGTAATTAGATAATTGTTGGGATTCCATTGTTGATAA
 AGGCTATAATATTAGGTATACAGAATATACTAGAAGTTCTCCTCGAGGAT
 TTAGGAATCCATAAAAGGGAATCTGCAATTCTACACAATTCTATAAATAT
 TATTATCATCGTTTTATATGTTAATATTCATTGATCCTATTACATTATCAAT

CCTTGCGTTTCAGCTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGT
CATCTTCTAACACCGTATATGATAATATACTAGTAACGTAAATACTAGTTA
GTAGATGATAGTTGATTTTTATTCCAACATACCACCCATAATGTAATAGAT
CTAAT

5

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 1559. This T2 control element has the DNA sequence

10 ATGAGATATATGTGGGTAATTAGATAATTGTTGGGATTCCATTGTTGATAA
AGGCTATAATATTAGGTATACAGAATATACTAGAAGTTCTCCTCGAGGAT
TTAGGAATCCATAAAAGGGAATCTGCAATTCTACACAATTCTATAAATAT
TATTATCATCGTTTTATATGTTAATATTCATTGATCCTATTACATTATCAAT
CCTTGCGTTTCAGCTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGT
CATCTTCTAACACCGTATATGATAATATACTAGTAACGTAAATACTAGTTA
15 GTAGATGATAGTTGATTTTTATTCCAACATACCACCCATAATGTAATAGAT
CTAAT

There are no genes controlled by this T1/T2 loop.

20

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

25

A C1/C2 short loop on chromosome 4 whose identifier is 1538 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

30

ATGAGATATATGTGGGTAATTAGATAATTGTTGGGATTCCATTGTTGATAA
AGGCTATAATATTAGGTATACAGAATATACTAGAAGTTCTCCTCGAGGAT
TTAGGAATCCATAAAAGGGAATCTGCAATTCTACACAATTCTATAAATAT
TATTATCATCGTTTTATATGTTAATATTCATTGATCCTATTACATTATCAAT
CCTTGCGTTTCAGCTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGT
CATCTTCTAACACCGTATATGATAATATACTAGTAACGTAAATACTAGTTA

GTAGATGATAGTTGATTTTTATTCCAACATACCACCCATAATGTAATAGAT
CTAATGAATCCATTTGTTTGTTAATAGTTT

This T1-T2 loop also modulates the C1/C2 short loops numbered 1539 to 1557

A C1/C2 short loop on chromosome 4 whose identifier is 1558 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

AGCTTCTCATAACTTATGTCATCATCTTAACACCGTATATGATAATATATT
GATAATATAACTTGTTGGAATAAAAAATCAACTATCATCTACTAACTAGTAT
TTACGTTACTAGTATATTATCATATACGGTGTTAGAAGATGACGCAAATG
ATGAGAAATAGTCATCTAAATTAGTGGAAGCTGA...GTCTATCTGGCGAAT
ATAAATTTTTACGCTACACACGTCATCGACATCTAAATATGACAGTCGCTG
AACTGTTCTTAGATATCCATGCTATTTATGAAGAACAACAGGGATCGAGA
AACAG

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 9 whose identifier is 3789 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YIL059C and has the DNA sequence

TTTATATGTTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCA
GCTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACA
CCGTATATGATAATATACTAGTAACGTAAATACTAGTTAGTAGATGATAG
TTGATTTTTATTCCAACAGTAT

The match between the T1 sequence and the C1/C2 sequence is

TTTATATGTTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCA
GCTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACA
CCGTATATGATAATATACTAGTAACGTAAATACTAGTTAGTAGATGATAG
TTGATTTTTATTCCAACA

5

The match between the T2 sequence and the C1/C2 sequence is

TTTATATGTTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCA
GCTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACA
CCGTATATGATAATATACTAGTAACGTAAATACTAGTTAGTAGATGATAG
TTGATTTTTATTCCAACA

10

A C1/C2 short loop on chromosome 12 whose identifier is 5289 controls the
expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
as a RNA single strand that is 3'UTR to the gene YLR301W and has the DNA
sequence

15

GGTGAATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATA
TTAGGTATGTAGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCAT
AAAAGGGAATCTGCAATTCTACACAATTCTATAAATATTATTATCATCGTT
TTATATGTTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAG
CTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACAC
CGTATATGATAATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGT
TGATTTTTATTCCAACAC

25

The match between the T1 sequence and the C1/C2 sequence is

AGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAA
TCTGCAATTCTACACAATTCTATAAATATTATTATCATCGTTTTATATGTTA
ATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAA
TTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTATATGAT

30

AATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTATT
CCAACA

The match between the T2 sequence and the C1/C2 sequence is

5

AGGATTTAGGAATCCATAAAAGGGAATCTGCAATTCTACACAATTCTATA
AATATTATTATCATCGTTTTATATGTTAATATTCATTGATCCTATTACATTA
TCAATCCTTGCGTTTCAGCTTCCACTAATTTAGATGACTATTTCTCATCATT
TGCATCATCTTCTAACACCGTATATGATAATATACTAGTAACGTAAATACT
10 AGTTAGTAGATGATAGTTGATTTTTATTCCAACA

A C1/C2 short loop on chromosome 13 whose identifier is 5753 controls the
expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed
as a RNA single strand that is 3'UTR to the gene YMR044W and has the DNA
15 sequence

15

TTGAGAAATGGGGGAATGTTGAGATAATTGTTGGGATTCCATTGTTGATA
AAGGCTATAATATTAGGTATACAGAATATACTAGAAGTTCTCCTCAAGGA
TATAGGAATCCTCAAAATGGAATCTATATTTCTACATACTAATATTACGAT
20 TATTCCTCATTCCGTTTTATATGTTTCATTATCCTATTACATTATCAATCCT
TGCACTTCAGCTTCCTCTAACTTCGATGACAGCTTCTCATAACTTATGTCA
TCATCTTAACACCGTATATGATAATATATTGATAATATAACTATTAGTTGA
TAGACGATAGTGGATTTTTATTCCAACAT

20

25 The match between the T1 sequence and the C1/C2 sequence is

AGATAATTGTTGGGATTCCATTGTTGATAAAGGCTATAATATTAGGTATAC
AGAATATACTAGAAGTTCTCCTC

30 The match between the T2 sequence and the C1/C2 sequence is

30

TTGTTGGGATTCCATTGTTGATAAAGGCTATAATATTAGGTATACAGAATA
TACTAGAAGTTCTCCTCAAGGAT

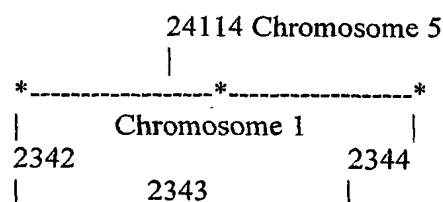
5

Two examples of multi-celled geneless connectrons – *C. elegans*

10

In the first example the existence of the T1-T2 (2342-2344) long loop is controlled by the C1/C2 (24114) short loop. The expression of one C1/C2 (2343) short loop is controlled by the existence of the T1-T2 (2342-2344) long loop.

15

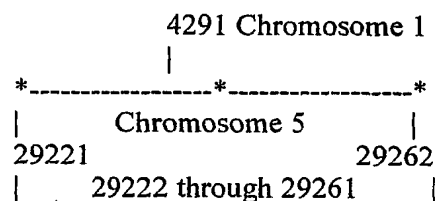


20

25

In the second example the existence of the T1-T2 (29221-29262) long loop is controlled by the C1/C2 (24114) short loop. The expression of one C1/C2 (2343) short loop is controlled by the existence of the T1-T2 (2342-2344) long loop.

30



35

A double stranded DNA loop of length 67.059 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 2342. This T1 control element has the DNA sequence

TGAAAACACTACAGTAATTCTTTAAATGACTACTGTAGC

5 This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 2344. This T2 control element has the DNA sequence

CTACTGTAGCGCTTGTGTCGATTTACGGGCTCGATT

10 There are no genes controlled by this T1/T2 loop.

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

15 A C1/C2 short loop on chromosome 1 whose identifier is 2343 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

TCGACACAAGCGCTACAGTAGCTATTTAAAGAATTACTGTAGTTTTCGCTA
CGAGATATTT

20 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

25 A C1/C2 short loop on chromosome 5 whose identifier is 24114 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene C13F10.5 and has the DNA sequence

30 GCGAAAACACTACAGTAATTCTTTAAATGACTACTGTAGCGCTTGTGTCGATT
TACGGGCTCGATTTTCG

The match between the T1 sequence and the C1/C2 sequence is

GAAAACTACAGTAATTCTTTAAATGACTACTGTAGC

The match between the T2 sequence and the C1/C2 sequence is

5

CTACTGTAGCGCTTGTGTCGATTTACGGGCTCGATT

10

A double stranded DNA loop of length 41.297 kilo-bases on chromosome 5 is bounded on the left by a T1 sequence whose identifier is 29221. This T1 control element has the DNA sequence

TTTAAATTTCCCGCCAAAAATTGACTGAAAATTTGGATTTTCTTTCCAAAA
ATTGACAGAAA

15

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 29262. This T2 control element has the DNA sequence

20

TGAAAATTTGAATTTCCCGCCAAAAATTAAC

There are no genes controlled by this T1/T2 loop.

25

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 5 whose identifier is 29222 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

30

AATTTCCCGCCAAAAATTGACTGAAAATTTGGATTTTCTTTCCAAAAATTG
ACAGAAA

This T1-T2 loop also modulates the C1/C2 short loops numbered 29223 to 29260

5 A C1/C2 short loop on chromosome 5 whose identifier is 29261 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

AAAATTGACTGAAAATTTGAATTTCCAGCCAAAAATTGACTGAAAATTTG
AATT

10 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

15 A C1/C2 short loop on chromosome 1 whose identifier is 4291 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene Y43F8C.5 and has the DNA sequence

20 AAAATTAAGTAAAATTTGAATTTCCCGCCAAAAATTGACTGAAAATTTG
AATTTCCCGCCAAAAAAATTGACTGAAAATTTGAATTTCCCGCCAAAA
TTGACTGAAAATTTGAATTTCCCGCCAAAAATTAATTGAAAATTTGAATTT
CCCGCCAAAAATTAATTGAACTTTGAATTTTCAA...ATTTCCCGCCAAAA
ATTAATTGAACTTTGAATTTTCAAATTTCCCGCCAAAAATTGACTGAAAA
TTTGAATTTCCCGCCAAAAATTAATTGAAAATTTGAATTTTGAATTTCCC
25 GCCAAAAATGACTGA

The match between the T1 sequence and the C1/C2 sequence is

TTTAAATTTCCCGCCAAAAATTGACTGAAAATTTG

30 The match between the T2 sequence and the C1/C2 sequence is

AAAAAAATTGACTGAAAATTTGAATTTCCCGCCAAAAATTGA

10. One connectron controls many geneless connectrons in single-celled and multi-celled eukaryotes

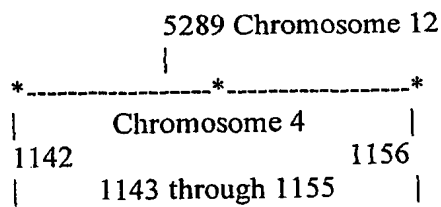
One C1/C2 short loop can control the existence of many geneless T1-T2 long loops.

5

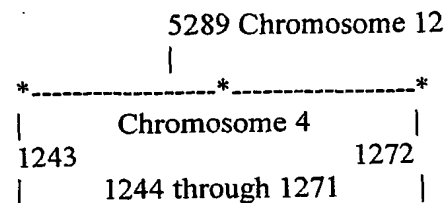
Example of a single-celled geneless connectron -- *S. cerevisiae*

In this example the existence of the three T1-T2 (1142-1156, 1242-1272 and 7102-7117) long loops is controlled by the C1/C2 (5289) short loop.

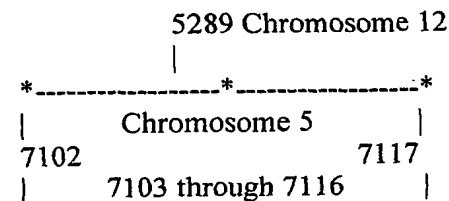
10



15



20



25

30

A double stranded DNA loop of length 5.337 kilo-bases on chromosome 4 is bounded on the left by a T1 sequence whose identifier is 1142. This T1 control element has the DNA sequence

35

ATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATATTAGG
TATGTAGATATACTAGAAGTTCTCCTCGAGGATTAGGAATCCATAAAAG

GGAATCTGCAATTCTACACAATTCTATAAATATTATTATCATCATTTTATA
TGTTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTCC
ACTAATTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTAT
ATGATAATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTT
5 TTATTCCAACA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 1156. This T2 control element has the DNA sequence

10 TTTTAATAAGGCAATAATATTAGGTATGTAGATATACTAGAAGTTCTCCTC
CAGGATTTAGGAATCCATAAAAAGGGAATCTGCAATTCTACACAATTCTAT
AAATATTATTATCATCATTTTATATGTTAATATTCATTGATCCTATTACATT
ATCAATCCTTGCGTTTCAGCTTCCACTAATTTAGATGACTATTTCTCATCAT
TTGCGTCATCTTCTAACACCGTATATGATAATATACTAGTAACGTAAATAC
15 TAGTTAGTAGATGATAGTTGATTTTTATTCCAACAAGAA

There are no genes controlled by this T1/T2 loop.

20 This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 4 whose identifier is 1143 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

25 ATTTTGAGATAATTGTTGGGATTCCATTTTTTAATAAGGCAATAATATTAGG
TATGTAGATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAAG
GGAATCTGCAATTCTACACAATTCTATAAATATTATTATCATCATTTTATA
TGTTAATATTCATTGATCCTATTACATTATCAAT...CTCTAAGTCTCATTGCC
30 TTTGTGCCAAAAAATCTGTTTCTAAATTTCTCTTCATTTGTAGACTTAATTA
TACTGATCGTTGATCTACTATCAGTAAGTAAGCCTTTAATAATTGGTTTCT
TGTTAAGTTCTTGACAAAGGTGACTGAGGTTATTCAATAGCGG

This T1-T2 loop also modulates the C1/C2 short loops numbered 1144 to 1154

5 A C1/C2 short loop on chromosome 4 whose identifier is 1155 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

GAGGAGAACTTCTAGTATATCTACATACCTAATATTATTGCCTTATTA
10 ATGGAATCCCAACAATTA

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

15 A C1/C2 short loop on chromosome 12 whose identifier is 5289 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YLR301W and has the DNA sequence

GGTGAATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATA
20 TTAGGTATGTAGAATATACTAGAAAGTTCTCCTCGAGGATTTAGGAATCCAT
AAAAGGGAATCTGCAATTCTACACAATTCTATAAATATTATTATCATCGTT
TTATATGTTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAG
CTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACAC
CGTATATGATAATATACTAGTACGTAAATACTAGTTAGTAGATGATAGTT
25 GATTTTATTCCAACAC

The match between the T1 sequence and the C1/C2 sequence is

ATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATATTAGG
30 TATGTAGA

The match between the T2 sequence and the C1/C2 sequence is

TTTTAATAAGGCAATAATATTAGGTATGTAGA

5

A double stranded DNA loop of length 5.251 kilo-bases on chromosome 4 is bounded on the left by a T1 sequence whose identifier is 1243. This T1 control element has the DNA sequence

10

CGTGTTTTATCTCATGTTGTTTCGTTTTGTTATTGAGATATATGTGGGTAATT
AGATAATTGTTGGGATTCCATTGTTGATAAAGGCTATAATATTAGGTATAC
AGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAA
TCTGCAATTCTACACAATTCTATAAAATATTATTATCATCGTTTTATATGTTA
ATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAA
15 TTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTATATGAT
AATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTATT
CCAACA

20

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 1272. This T2 control element has the DNA sequence

25

TGAGATATATGTGGGTAATTAGATAATTGTTGGGATTCCATTGTTGATAAA
GGCTATAATATTAGGTATACAGAATATACTAGAAGTTCTCCTCGAGGATTT
AGGAATCCATAAAAGGGAATCTGCAATTCTACACAATTCTATAAAATATTA
TTATCATCGTTTTATATGTTAATATTCATTGATC...TATACTAGTAACGTAA
ATACTAGTTAGTAGATGATAGTTGATTTTTATTCCAACAGTTATAAGGTTG
TTTCATATGTGTTTTATGAA

30

There are no genes controlled by this T1/T2 loop.

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 4 whose identifier is 1244 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

5

TTTATCTCATGTTGTTTCGTTTTGTTATTGAGATATATGTGGGTAATTAGATA
ATTGTTGGGATTCCATTGTTGATAAAGGCTATAATATTAGGTATACAGAAT
ATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAATCTGC
AATTCTACACAATTCTATAAATATTATTATCAT...GTCTCGATGTAGTATAC
10 GTATAAATTATTACCTGATACTTCATCTCTAAGTCTCATTGCCTTTGTGCCA
AAAAATCTGTTTCTAAATTTCTCTTCATTTGTAGACTTAATTATACTGATCG
TTGATCTACTATCAGTAAGT

This T1-T2 loop also modulates the C1/C2 short loops numbered 1245 to 1270

15

A C1/C2 short loop on chromosome 4 whose identifier is 1271 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

20

TGTTGTATCTCAAAATGAGATATGTCAGTATGACAATACGTCATCCTAAAC
GTTCATAAAACACATATGAAACAACCTTATAACTGTTGGAATAAAAATCA
ACTATCATCTACTAACTAGTATTTACGTTACTAGTATATTATCATATACGG
TGTTAGAAGATGACGCAAATGATGAGAAATAGTC...CAACAATGGAATCC
CAACAATTATCTAATTACCCACATATATCTCATGGTAGCGCCTGTGCTTCG
25 GTTACTTCTAAGGAAGTCCACACAAATCAAGATCCGTTAGACGTTTCAGC
TTCCAAAA

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

30

A C1/C2 short loop on chromosome 12 whose identifier is 5289 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed

as a RNA single strand that is 3'UTR to the gene YLR301W and has the DNA sequence

5 GGTGAATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATA
TTAGGTATGTAGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCAT
AAAAGGGAATCTGCAATTCTACACAATTCTATAAATATTATTATCATCGTT
TTATATGTTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAG
CTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACAC
CGTATATGATAATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGT
10 TGATTTTTATTCCAACAC

The match between the T1 sequence and the C1/C2 sequence is

15 AGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAA
TCTGCAATTCTACACAATTCTATAAATATTATTATCATCGTTTTATATGTTA
ATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAA
TTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTATATGAT
AATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTATT
CCAACA

20

The match between the T2 sequence and the C1/C2 sequence is

25 AGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAA
TCTGCAATTCTACACAATTCTATAAATATTATTATCATCGTTTTATATGTTA
ATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAA
TTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTATATGAT
AATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTATT
CCAACA

30

A double stranded DNA loop of length 5.296 kilo-bases on chromosome 15 is bounded on the left by a T1 sequence whose identifier is 7102. This T1 control element has the DNA sequence

5

CATGATTAATATGACCAATCGGCGTGTGTTTTTGAAAAGTGGGTGAATTTT
GAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATATTAGGTATGT
AGAATGTACTAGAAGTTCTCCTCAAGGATTTAGGAATCCATGAAAGGGAA
TCTGCAATTCTACACAATTCTATAAATATTATTATCATCATTTTATATGTTA
10 ATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAA
TTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTATATGAT
AATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTATT
CCAACA

15

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 7117. This T2 control element has the DNA sequence

20

TGAAAAGTGGGTGAATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAG
GCAATAATATTAGGTATGTAGAATGTACTAGAAGTTCTCCTCAAGGATTT
AGGAATCCATGAAAGGGAATCTGCAATTCTACACAATTCTATAAATATTA
TTATCATCATTTTATATGTTAATATTCATTGATCCTATTACATTATCAATCC
TTGCGTTTCAGCTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGTCA
TCTTCTAACACCGTATATGATAATATACTAGTAACGTAAATACTAGTTAGT
25 AGATGATAGTTGATTTTTATTCCAACAGTTTTATATACCTCTCTTATTTAGT
ATAAGAA

There are no genes controlled by this T1/T2 loop.

30

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 15 whose identifier is 7103 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

5 AAGAACATTGCTGATGTGATGACAAAACCTCTTCCGATAAAAACATTTAA
ACTATTA ACTAACA AATGGATT CATTAGATCTATTACATTATGGGTGGTAT
GTTGGAATAAAAATCAACTATCATCTACTAACTAGTATTTACGTTACTAGT
ATATTATCATATACGGTGTTAGAAGATGACGCAAATGATGAGAAATAGTC
ATCTAAATTAGTGGAAGCTGAAACGCAAGGATTGATAATGTAATAGGATC
10 AATGAATATTAACATATAAAATGATGATAATAATATTTATAGAATTGTGT
AGAATTGCAGATTCCCTTTTCATGGATTCCCTAAATCCTTGAGGAGAACTTCT
AGTA

This T1-T2 loop also modulates the C1/C2 short loops numbered 7104 to 7115

15

A C1/C2 short loop on chromosome 15 whose identifier is 7116 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

20 CCATTCTGTGGAGGTGGTACTGAAGCAGGTTGAGGAGAGACATGATGATG
GTTCTCTGGAACAGCT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

25

A C1/C2 short loop on chromosome 12 whose identifier is 5289 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YLR301W and has the DNA sequence

30

GGTGAATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATA
TTAGGTATGTAGAATATACTAGAAGTTCTCCTCGAGGATTAGGAATCCAT

.AAAAGGGAATCTGCAATTCTACACAATTCTATAAATATTATTATCATCGTT
 TTATATGTTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAG
 CTTCCACTAATTTAGATGACTATTTCTCATCTTTGCGTCATCTTCTAACAC
 CGTATATGATAATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGT
 5 TGATTTTATTCCAACAC

The match between the T1 sequence and the C1/C2 sequence is

10 GGTGAATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATA
 TTAGGTATGTAGAAT

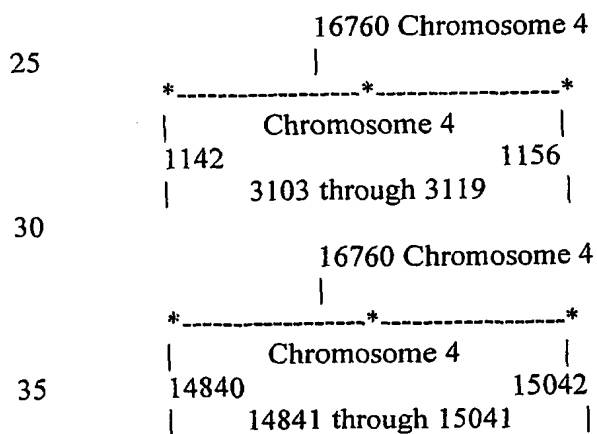
The match between the T2 sequence and the C1/C2 sequence is

15 GGTGAATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATA
 TTAGGTATGTAGAAT

Example of a multi-celled geneless connectron – *C. elegans*

20

In this example the existence of the three T1-T2 (1142-1156, 14840-15042 and 15365-15627) long loops is controlled by the C1/C2 (16760) short loop.



16760 Chromosome 4
----------*
5 | Chromosome 5 |
15365 15627
| 15366 through 15625 |

10

A double stranded DNA loop of length 15.894 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 3101. This T1 control element has the DNA sequence

15

CAAATCGGCAAATTGCCGGAATTGAACATTTC

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 3120. This T2 control element has the DNA sequence

20

AAACGATTTTCCGGCAAATCGGCAAATTGCCGGAATTGTAATTTCCGGC
AAAT

There are no genes controlled by this T1/T2 loop.

25

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 1 whose identifier is 3103 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

30

TTAAAATTTCCGGCAAATCGGCAAATTGGCAGAAATGAACTCACGGCAA
ATCGG

35

This T1-T2 loop also modulates the C1/C2 short loops numbered 3104 to 3118

A C1/C2 short loop on chromosome 1 whose identifier is 3119 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

5
CCCGCATTTTTGTAGATCAAACCGTAATGGGACGGCCTGGCAACACGTG
ATTTTCCAAAT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

10
A C1/C2 short loop on chromosome 4 whose identifier is 16760 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene T23E1.2 and has the DNA sequence

15
GGCAAATTGCCGAAATTGAACATTTCCGGCAAATCGGCAAATTGCCGGAA
TTGAACATTTCCGGCAAATCGGCAAATTGCCGGAATTGAACATTTCCGGC
AAATCGGCAAATTGCCGGAATTGA

20 The match between the T1 sequence and the C1/C2 sequence is

CAAATCGGCAAATTGCCGGAATTGAACATTTCC

The match between the T2 sequence and the C1/C2 sequence is

25
TTTCCGGCAAATCGGCAAATTGCCGGAATTG

30 A double stranded DNA loop of length 86.977 kilo-bases on chromosome 3 is bounded on the left by a T1 sequence whose identifier is 14840. This T1 control element has the DNA sequence

AAAAATTTCCGGCAAGTCGGCAATTTTCCGAAAATGAAAATTTCCGGCAA
ATCGGCAAATTGCCGGAATTGAAAATTCCTGGCAAATCAGCAAATTTGCG
GCAAATCGGCAATTTGCCGAAAATGAAAATTTCCGGCAAAT

5

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 15042. This T2 control element has the DNA sequence

10

CAAATCGGTAGGTAAATTGGCCAACTTGAAAATTTCCGGCAAATCGGCA
AATTCGCGAACTGAACATTTCCGGCAAATCGGCAAATTGCTCGAACT

There are no genes controlled by this T1/T2 loop.

15

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 3 whose identifier is 14841 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

20

AAAAATTTCCGGCAAGTCGGCAATTTTCCGAAAATGAAAATTTCCGGCAA
ATCGGCAAATTGCCGGAATTGAAAATTCCTGGCAAATCAGCAAATTTGCG
GCAAATCGGCAATTTGCCGAAAATGAAAATTTCCGGCAAAT

25

This T1-T2 loop also modulates the C1/C2 short loops numbered 14842 to 15040

A C1/C2 short loop on chromosome 3 whose identifier is 15041 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

30

CGGCAATTGCCGTTCCGGCAATTTGCCAATTGCCGGAATTTCAATTCCG
GCAA

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

5 A C1/C2 short loop on chromosome 4 whose identifier is 16760 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene T23E1.2 and has the DNA sequence

10 GGCAAATTGCCGAAATTGAACATTTCCGGCAAATCGGCAAATTGCCGGAA
TTGAACATTTCCGGCAAATCGGCAAATTGCCGGAATTGAACATTTCCGGC
AAATCGGCAAATTGCCGGAATTGA

The match between the T1 sequence and the C1/C2 sequence is

15 ATTTCCGGCAAATCGGCAAATTGCCGGAATTGAA

The match between the T2 sequence and the C1/C2 sequence is

20 TGAACATTTCCGGCAAATCGGCAAATTGC

A double stranded DNA loop of length 98.488 kilo-bases on chromosome 3 is bounded on the left by a T1 sequence whose identifier is 15365. This T1 control
25 element has the DNA sequence

30 AAAATTTCCGGCAAATCGGCAATTTGCCAAAAATTGAAATTTCCGGCAA
TCGGCAATTTGTCAAAAATGAAAATTTCCGGCAAATCGGCAAATTGCCGA
AAATGAAAATTTCCGGCAAATCGGCAAACCTCCGGAAGTAAAATTTCCG
GCAAATCGGCAATTTGCCATAAATGAACATTTCCGG...GGCGAAAATTAAA
ATTTCCGCCATATCGGCAATTTGCCAAAAAATTAAAATTTCCGGCAAATC

GGCAAATTGCCGGAATTCAAAATTTCCGGCAAACCGGCAAATTGCCGGAA
CTCAAATTTCCCGGCAAATCAGCAAATTGCCGGAATT

5 This double stranded DNA loop is bounded on the right by a T2 control element
whose identifier is 15627. This T2 control element has the DNA sequence

TGGCAAACCGGCAAATTGCCGGAATTGAACATTTCCGGCAAATCGGCAAT
TTGCCGGAATTGAAATTT

10 There are no genes controlled by this T1/T2 loop.

This long T1/T2 double stranded DNA loop modulates the expression of the
following C1/C2 short loops

15 A C1/C2 short loop on chromosome 3 whose identifier is 15366 controls the
expression of the genes of one or more other T1/T2 long loops. This C1/C2 short
loop has the DNA sequence

20 TGCCGATTTGCCGGAATTTTCATTTTCGGCAATTTGCCGATTTGCCGGAA
ATTTTCATT

This T1-T2 loop also modulates the C1/C2 short loops numbered 15366 to 15624

25 A C1/C2 short loop on chromosome 3 whose identifier is 15625 controls the
expression of the genes of one or more other T1/T2 long loops. This C1/C2 short
loop has the DNA sequence

30 TCAAGCAAATTGTCAAATTCGCGGAACTAAACATTTCCGGCAAATCGGCA
AATT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2
short loops.

A C1/C2 short loop on chromosome 4 whose identifier is 16760 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene T23E1.2 and has the DNA sequence

5

GGCAAATTGCCGAAATTGAACATTTCCGGCAAATCGGCAAATTGCCGGAA
TTGAACATTTCCGGCAAATCGGCAAATTGCCGGAATTGAACATTTCCGGC
AAATCGGCAAATTGCCGGAATTGA

10

The match between the T1 sequence and the C1/C2 sequence is

ATTTCCGGCAAATCGGCAAATTGCCGGAATT

The match between the T2 sequence and the C1/C2 sequence is

15

CGGCAAATTGCCGGAATTGAACATTTCCGGCAAATCGGCAA

Claims

What is claimed is:

1. A method of identifying DNA sequences that control the expression of different collections of genes in a genome comprising, detecting selected DNA sequences adjacent to some genes excluding exons and introns.
2. A method of identifying DNA sequences that control the expression of different collections of genes comprising, detecting, by computer, one or more pairs of non-adjacent DNA sequences to which are bound to two RNA sequences.
3. A method of identifying DNA sequences that control the expression of different collections of genes in a genome comprising detecting changes in connectron behavior in the genome.
4. A method of modifying the expression of different gene collections in a genome, comprising detecting changes in connectron behavior as a result of an exogenous stimulus.
5. A method of detecting where and when new genes are being integrated into a host genome comprising detecting the connectrons in said host genome.
6. A method of detecting the expression effect of different gene collections in a given body comprising detecting the back and forth flow of connectrons between the chromosomes thereof.
7. A method of modifying a given body comprising modifying the connectron organization therein.

8. A method of detecting connectron control and target sequences in a given genome comprising:

determining the base composition of said genome,
determining one or more sites of control sequence organization, and/or
determining one or more sites of target application.

9. A method of determining the response of a cell in any tissue to changes in the cell's environment and/or genetic composition comprising providing a complete genomic DNA sequence for the organism and determining the effect of changes in connectrons due to application of a given exogenous stimulus to the genome.

10. In prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes, the tetradic relationship $T1=C1$ and $T2=C2$ where T1 and T2 are DNA sequences 20 or more bases in length, where the C1 sequence is adjacent to the C2 sequence, where the T1 and T2 sequences are on the same chromosome, and where the C1/C2 sequences are on the same chromosome as T1 and T2 or where the C1/C2 sequences are on a chromosome different from T1 and T2, wherein:

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

11. In prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes, the connectron relationship that permits many different C1/C2 short loops to control the existence of a T1-T2 long loop and wherein said C1/C2 short loops can be on the same chromosome or on different chromosomes from the T1-T2 long loop, wherein:

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

C1/C2 - any positive or negative strand DNA sequence of 540 or more bases such that the C1 sequence is adjacent to the C2 sequence,

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

12. In prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes, the connectron relationship that permits one C1/C2 short loop to control the existence of many T1-T2 long loops, the C1/C2 short loop can be on the same chromosome or on different chromosomes from the T1-T2 long loops, wherein:

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

13. The connectron relationships between prokaryotes and their plasmids wherein said connectrons implement a control mechanism between the two genomes that makes it possible from them to form a symbiotic relationship, and in the case of *D. radiodurans* the relationship is not symmetric, and the *D. radiodurans* genome sends C1/C2 short loops to the MP1 plasmid, wherein:

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

14. The connectron relationships that exist in plant and higher animals.
15. In prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes, the connectron relationship that permits one C1/C2 short loop to control the existence of one or more T1-T2 long loops without being subject to any expression controls other than those of the gene to which the C1/C2 is 3'UTR, wherein:

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart,

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart, and

3'UTR - untranslated 3' end of an mRNA is beyond the end of the last exon, a stop codon in the mRNA causes the ribosome to stop the translation of mRNA into protein.

16. In prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes, the connectron relationship that permits one C1/C2 short loop to control the existence of one or more T1-T2 long loops such that this C1/C2 short loop is itself subject to expression control by another T1-T2 long loop which surrounds it, wherein:

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

17. In prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes, the connectron relationship that permits one C1/C2 short loop to control the existence of the T1-T2 long loop that surrounds it, wherein:

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

C1/C2 - any positive or negative strand DNA sequence of 50 or more bases such that the C1 sequence is adjacent to the C2 sequence,

T1 sequence - any positive or negative strand DNA sequence of 20bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

18. The connectron relationship that do not have any genes within the T1-T2 long loop, wherein:

T1 sequence is any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, and

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, and the T2 or T1 sequences must be between about 1kb and 105kb apart.

19. The geneless connectron relationship where one C1/C2 short loop controls the existence of many geneless T1-T2 long loops, wherein:

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

1/26

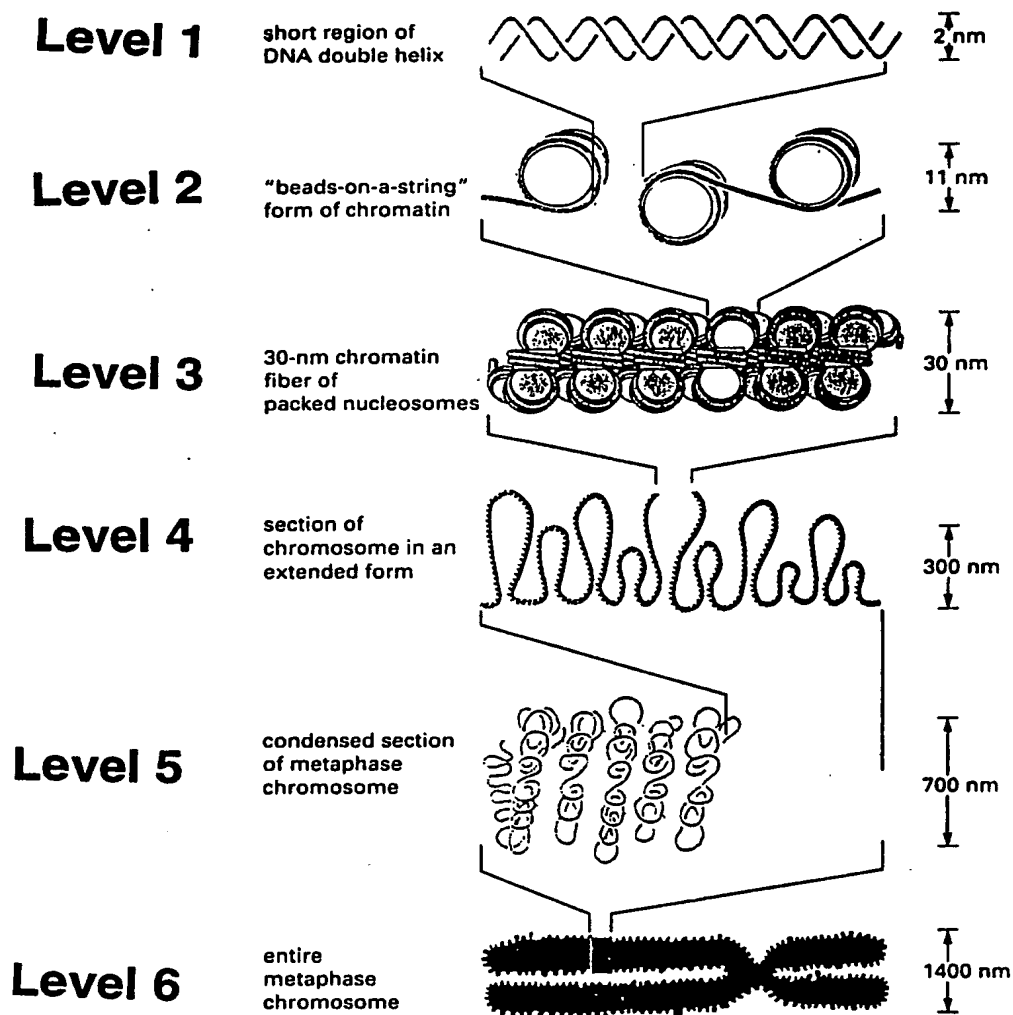
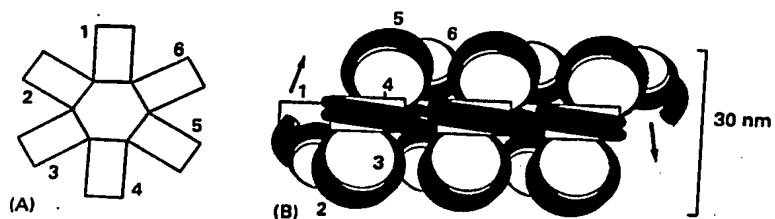
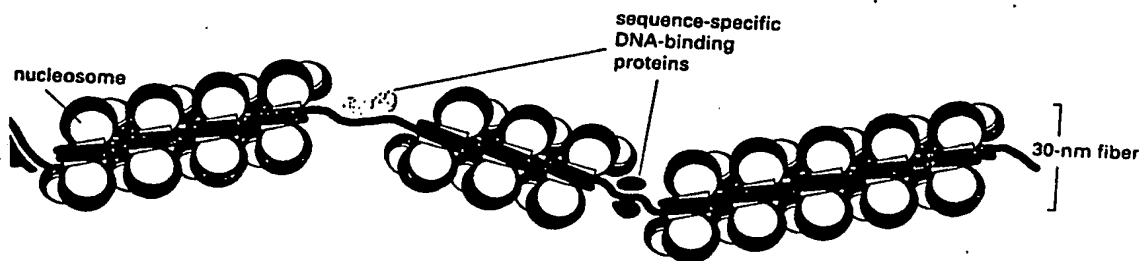


Figure 1

2/26



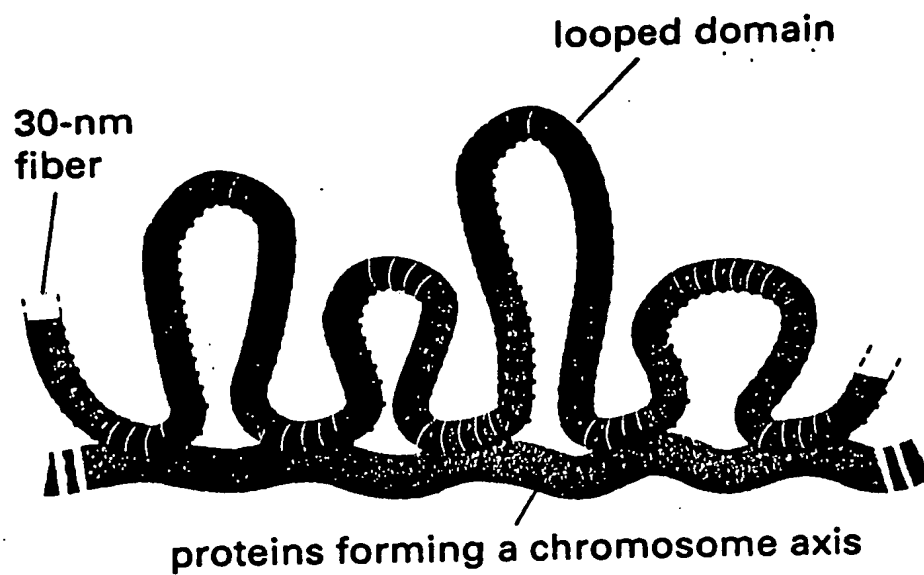
(a)



(b)

Figure 2

3/26

**Figure 3**

4/26

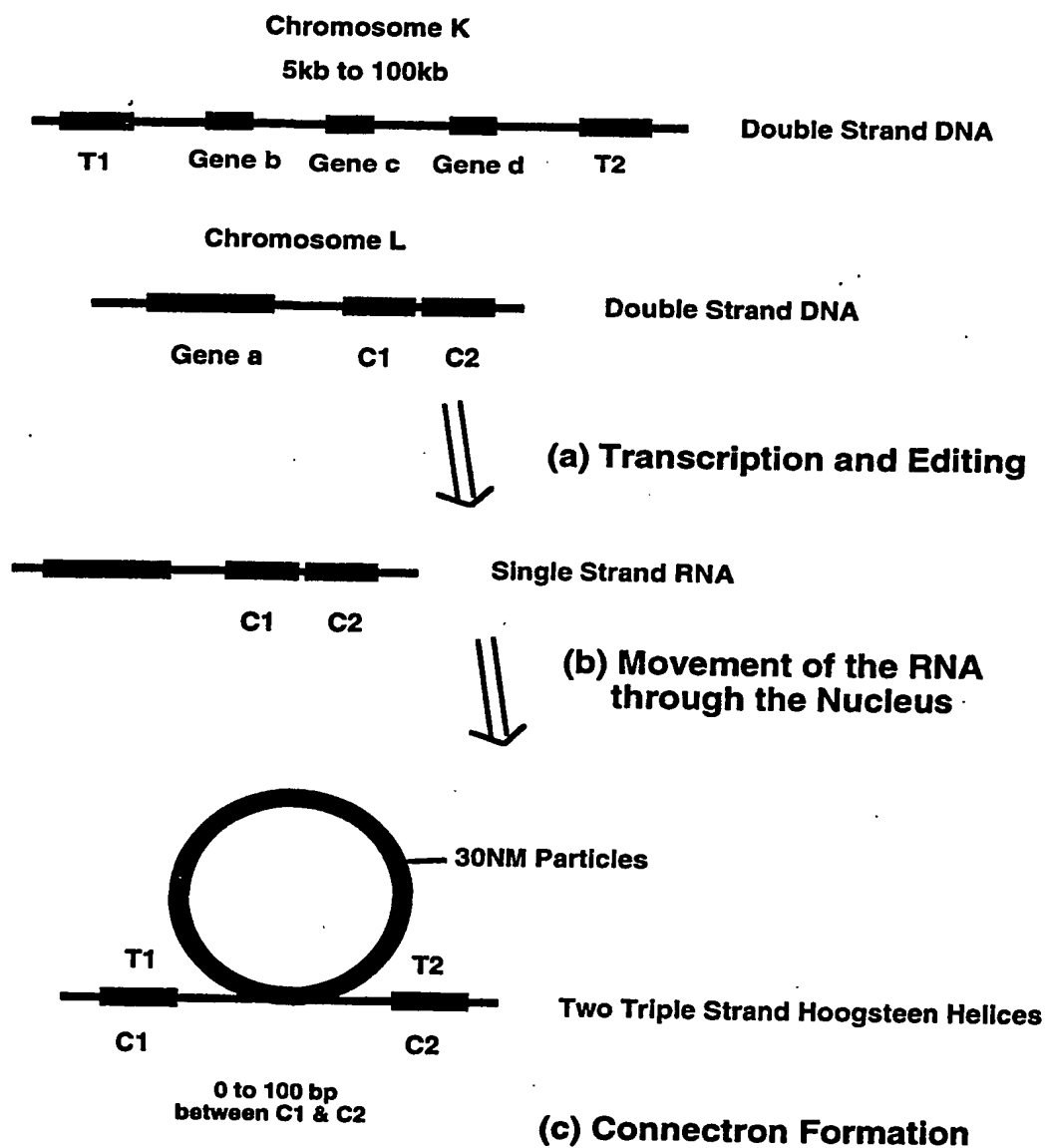


Figure 4

5/26

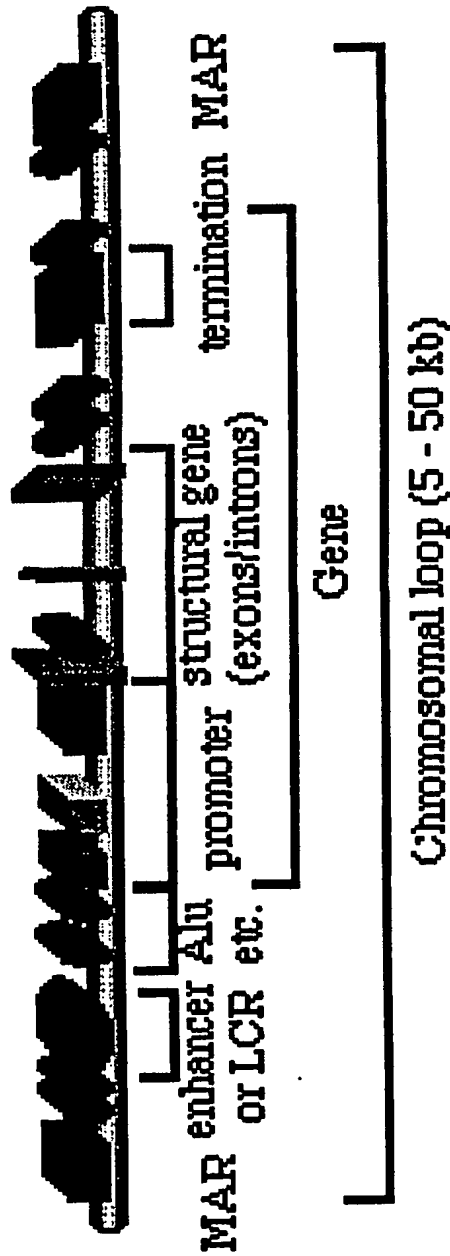


Figure 5

6/26

Genome	Chromosome Number	Number of Genes	Chromosome Length (bp)	Total Number of C1/C2 Short Loops	Total Number of T1/T2 Long Loops	Number of Unique C1/C2 Short Loops	Number of Unique T1/T2 Long Loops	Percent of Genes Controlled
				544	544	214	178	24.78
<i>E. coli</i>	I	4,289	4,639,221					
<i>H. pylori</i>	I	1,566	1,667,867	19	19	17	10	10.53
<i>M. luteus</i>	I	1,715	1,664,970	1,848	1,648	122	360	35.12
<i>D. radiodurans</i>	I	2,580	2,648,638	974	628	246	292	
	II	357	412,348	113	14	17	10	
III - Plasmid MP1		131	177,466	903	1,259	87	464	
IV - Plasmid CP1		35	45,704	91	80	28	45	77.14
Totals		3,103	3,284,156	1,981	1,981	378	811	
<i>S. cerevisiae</i>	I	110	230,209	898	283	26	112	
	II	447	813,137	1,387	2,516	568	1,094	
	III	172	315,339	1,314	1,762	336	553	
	IV	845	1,520,519	4,253	3,689	1,298	1,278	
	V	280	576,970	2,034	2,888	638	1,111	
	VI	127	270,148	438	155	153	45	
	VII	629	1,090,936	3,003	2,776	981	982	
	VIII	278	562,638	1,207	410	433	166	
	IX	236	439,885	248	81	113	48	
	X	406	745,440	1,691	2,019	541	768	
	XI	351	666,448	513	228	128	68	
	XII	578	1,078,173	3,010	4,282	951	1,921	
	XIII	501	924,430	2,306	3,254	680	1,229	
	XIV	458	784,328	1,579	490	528	144	
	XV	609	1,091,283	1,630	626	543	210	
	XVI	510	948,066	2,628	2,680	820	967	
Totals		6,538	12,057,849	28,139	28,139	8,737	10,697	19.88
<i>C. elegans</i>	I	2,474	16,183,833	315,926	159,815	12,245	24,537	
	II	3,043	17,004,825	133,735	92,585	13,370	26,647	
	III	2,100	12,114,540	93,171	897,303	8,472	138,469	
	IV	2,523	15,887,371	67,196	99,562	10,641	14,420	
	V	4,358	21,280,512	337,841	246,674	15,612	33,571	
	X	2,625	17,624,844	611,789	63,719	6,138	7,805	
Totals		17,123	100,096,025	1,559,658	1,559,658	66,478	245,449	33.6

Table 1 - page 1

7/26

Genome	Chromosome Number	Number of Genes	Chromosome Length (bp)	Total Number of C1/C2 Short Loops	Total Number of T1/T2 Long Loops	Number of Unique C1/C2 Short Loops	Number of Unique T1/T2 Long Loops	Percent of Genes Controlled
	Scaffold							
D. melanogaster	1	1,138	10,564,115	16,200	17,916	3,268	5,015	
	2	698	4,515,887	3,293	9,317	955	2,634	
	3	633	5,077,262	3,502	3,027	1,038	558	
	4	1,075	9,080,116	129,140	127,538	1,566	5,910	
	5	1,777	15,101,630	16,800	16,463	2,565	2,093	
	6	122	1,265,848	668	237	190	110	
	7	28	536,527	476	586	208	378	
	8	615	4,861,285	10,727	10,731	603	978	
	9	317	2,892,893	2,506	759	624	214	
	10	286	2,159,491	1,436	1,049	494	369	
	11	1,822	16,346,801	8,593	8,415	2,363	1,860	
	12	143	1,363,555	151	15	64	13	
	13	3,075	24,259,494	11,273	9,274	3,133	1,608	
	14	49	705,960	1,922	1,754	606	662	
	15	91	1,361,891	3,544	4,719	907	2,721	
	16	66	509,295	101	106	42	28	
	17	1,893	13,672,780	6,770	3,286	1,929	739	
	18	61	751,204	1,451	2,521	490	1,596	
	19	35	403,964	471	1,291	212	926	
	Totals	13,924	115,228,998	219,014	219,014	21,257	28,411	5.21
A. thaliana	a	3,221	14,221,815	22,735	20,165	3,310	4,603	
	b	3,136	14,668,893	22,202	25,016	4,043	7,324	
	c	4,036	19,647,091	153,505	75,591	10,377	19,290	
	IV	3,823	17,550,082	523,824	601,594	6,445	54,956	
	Totals	14,216	66,087,871	722,366	722,366	24,175	86,173	7.51
H. sapiens	XX	position data not available	33,667,731		> 10,000			

Table 1 - page 2

8/26

Chromosome of C1/C2	Chromosome of T1-T2															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	39	57	35	57	4	5	19	2	41		38	25	7	12	12	
2	9	70	13	27	98		22	17	1		9	36	130	15		89
3	6		54	44	39		37	5	1		5	16	23	10	6	29
4	10	73	93	257	158	13	138	12	2	67	9	135	144	30	38	125
5	1	47	13	42	32	5	19		1	26		76	14	5	9	18
6																46
7	4	72	46	114	83	7	100	14	2	64	6	83	91	15	20	73
8	10	66	13	52			25	14	3	25		58	15	3	12	16
9	2			12	14		18	9	2	1		4		2	4	8
10	4	9	24	48	107	1	47	9	1	59	8	38	100	10	8	81
11		27	15	39	1		15			26	5	6	29	1	11	14
12	10	68	61	107	90	8	77	2		56	2	193	73	19	25	76
13	4	78	60	134	85	2	68	8	3	71	14	60	173	10	28	127
14		66	13	56	14		16			25		15	18	4	12	28
15	3	10	20	31	79	1	27	11	1	8	8	52	78	6	7	82
16	6	54	68	126	121	9	86	14	1	51	15	100	130	20	24	137

Table 2

9/26

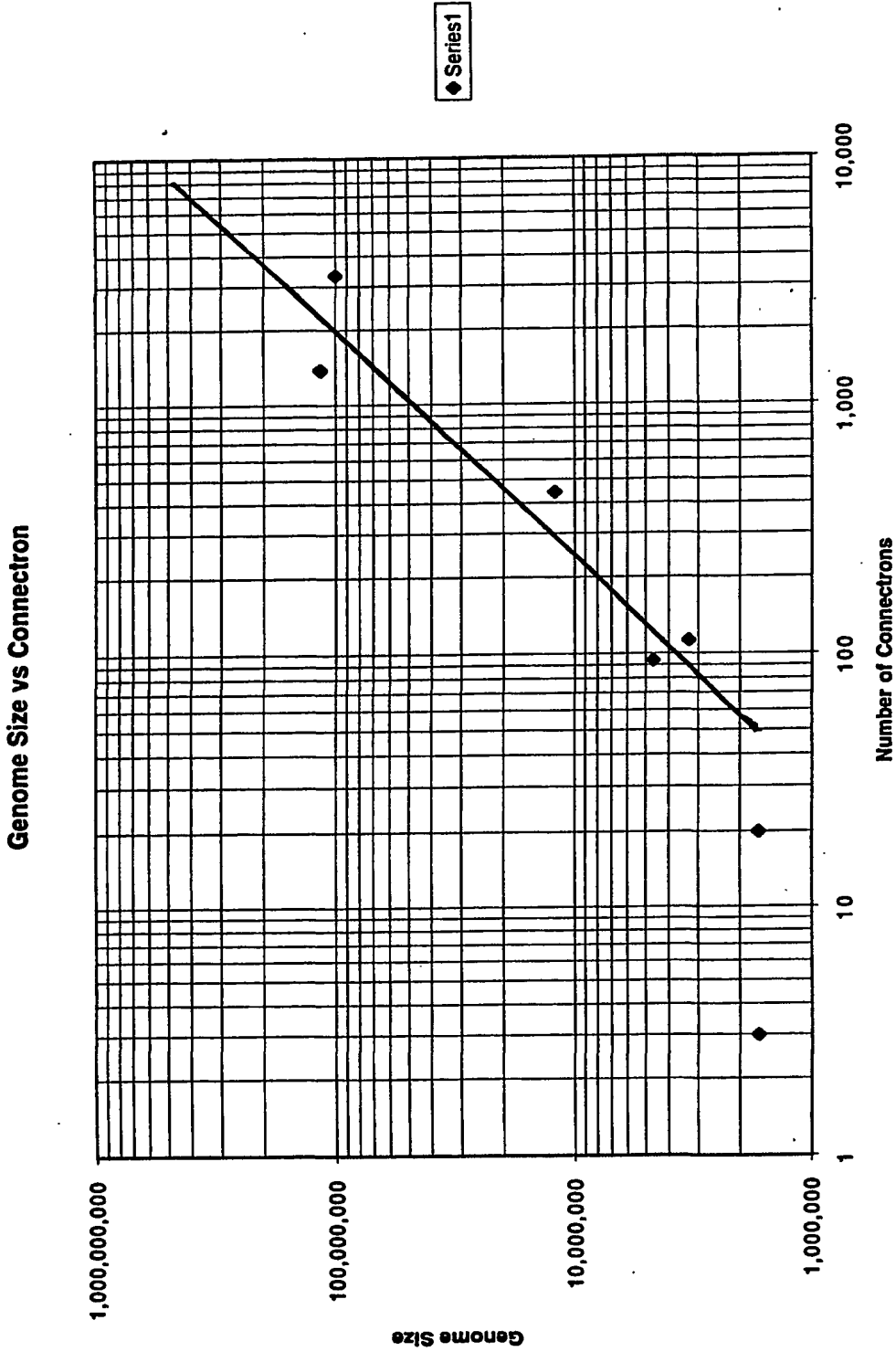


Figure 6

10/26

S. Cervesiae - Fragment Distribution

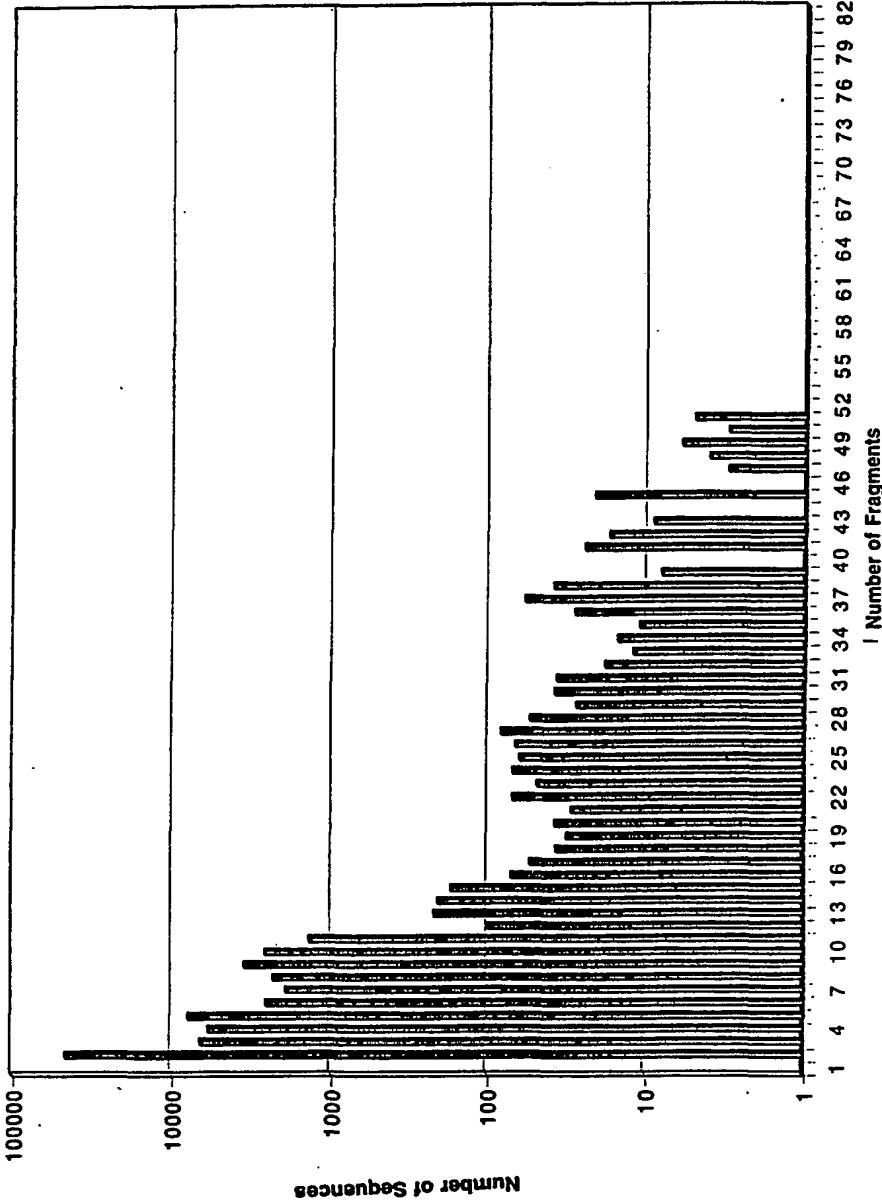
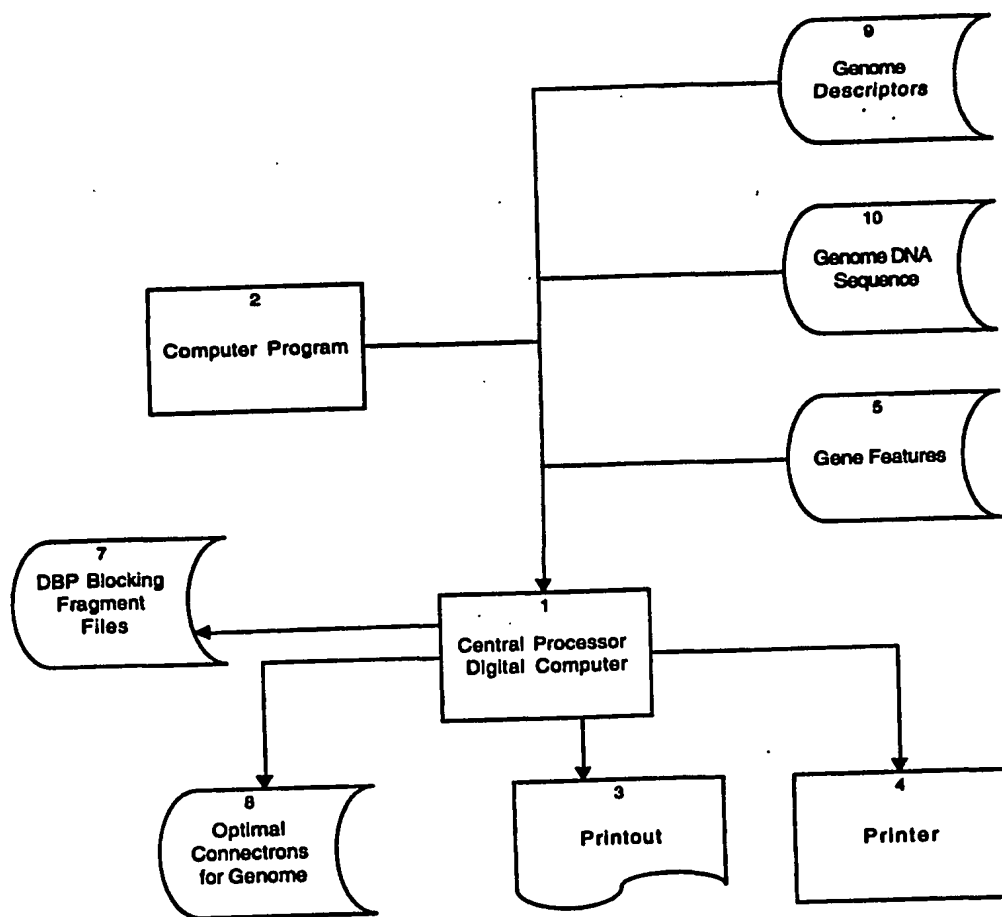
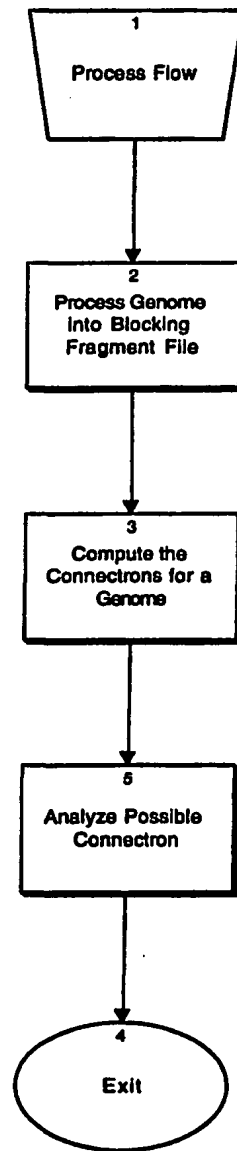


Figure 7

11/26

**Figure 8**

12/26**Figure 9**

13/26

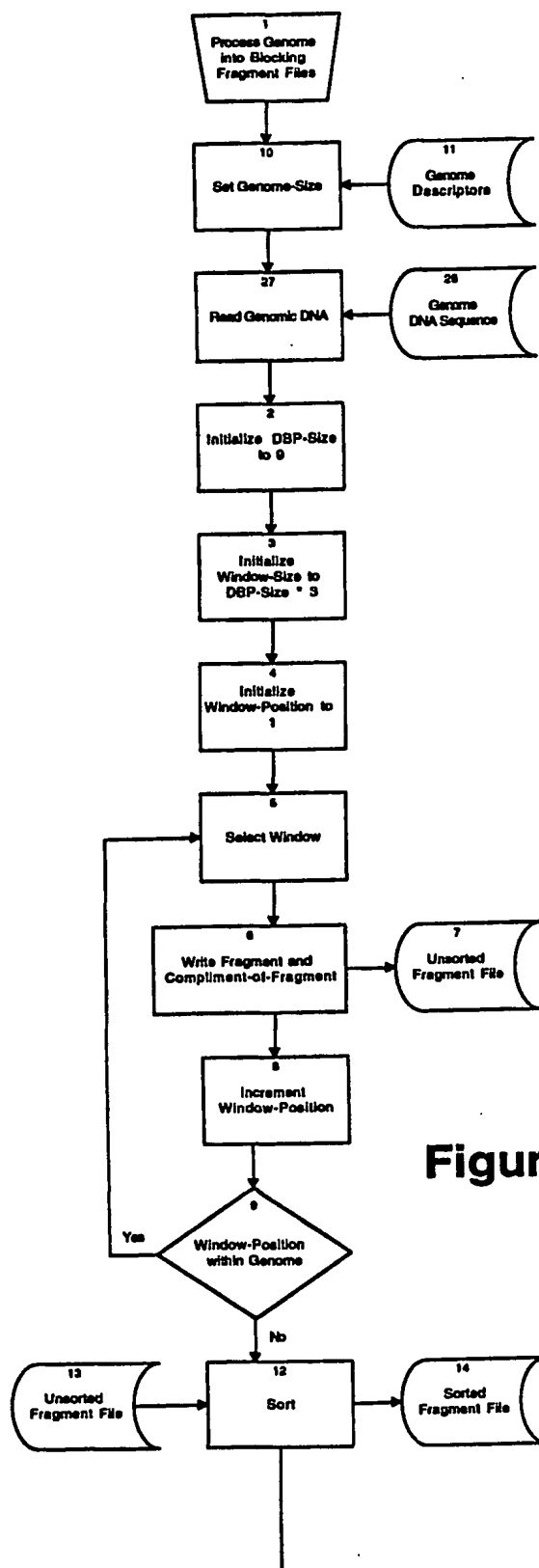


Figure 10 - page 1

14/26

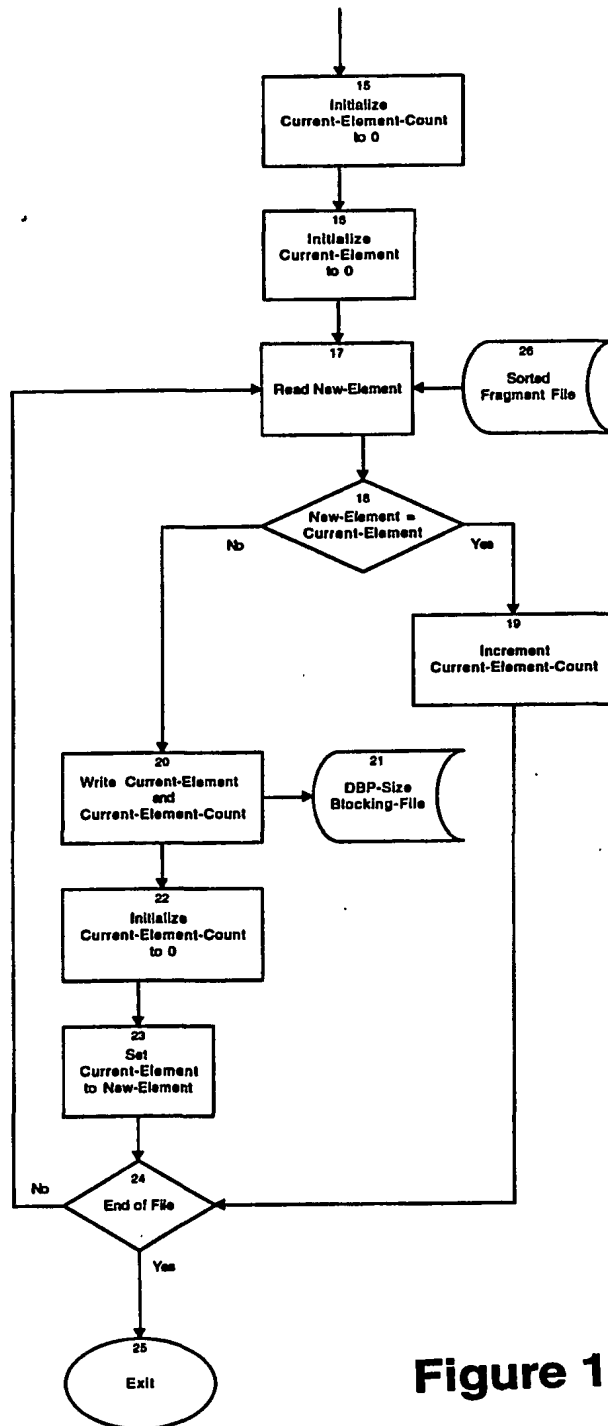
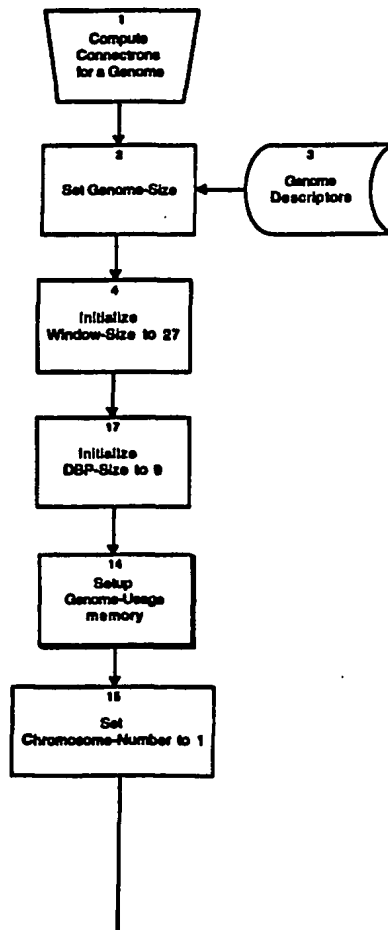


Figure 10 - page 2

15/26

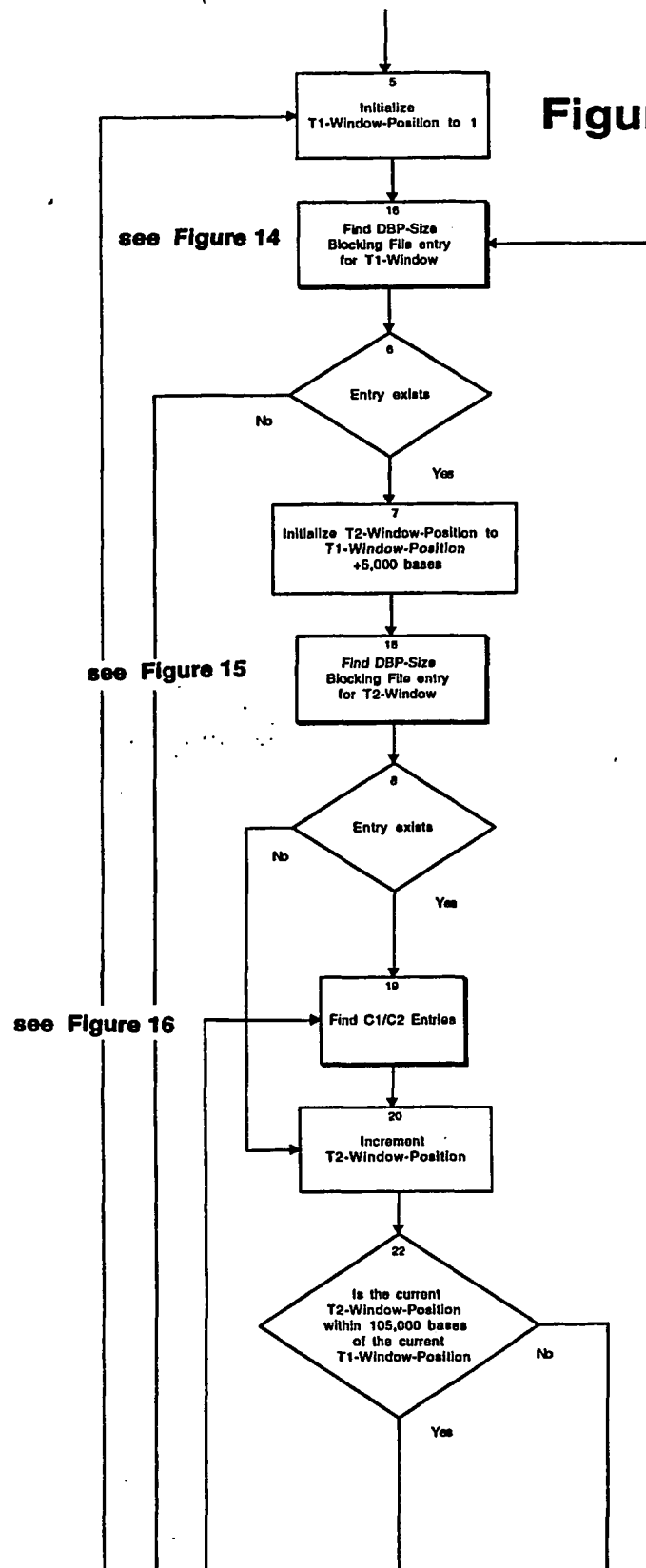
Figure 11 - page 1

see Figure 13



16/26

Figure 11 - page 2



17/26

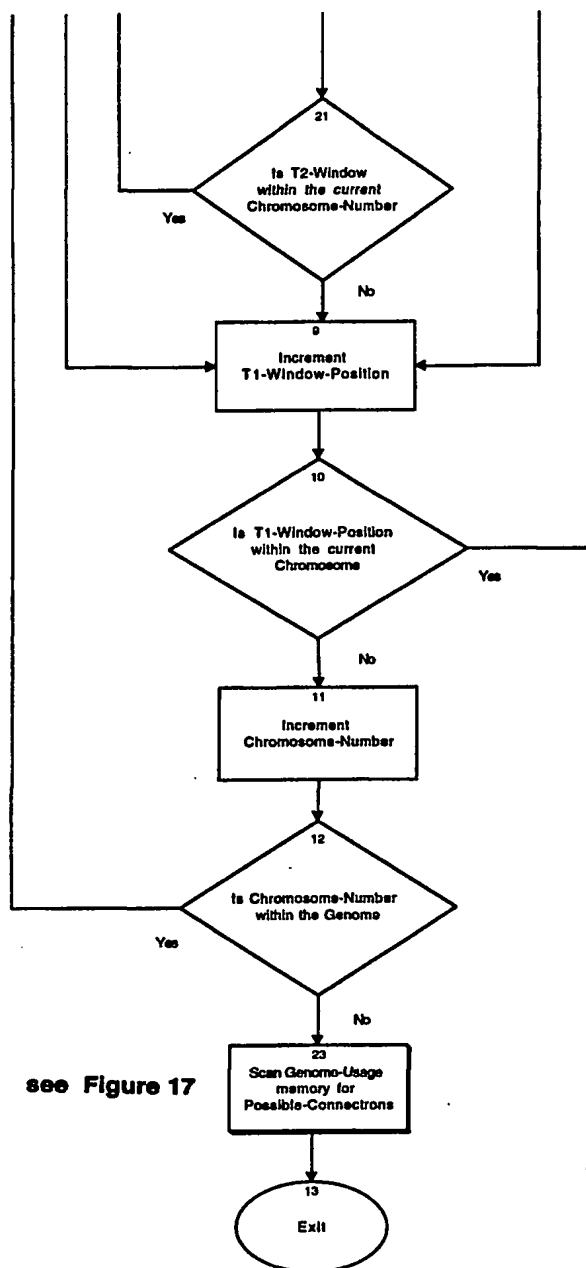


Figure 11 - page 3

18/26

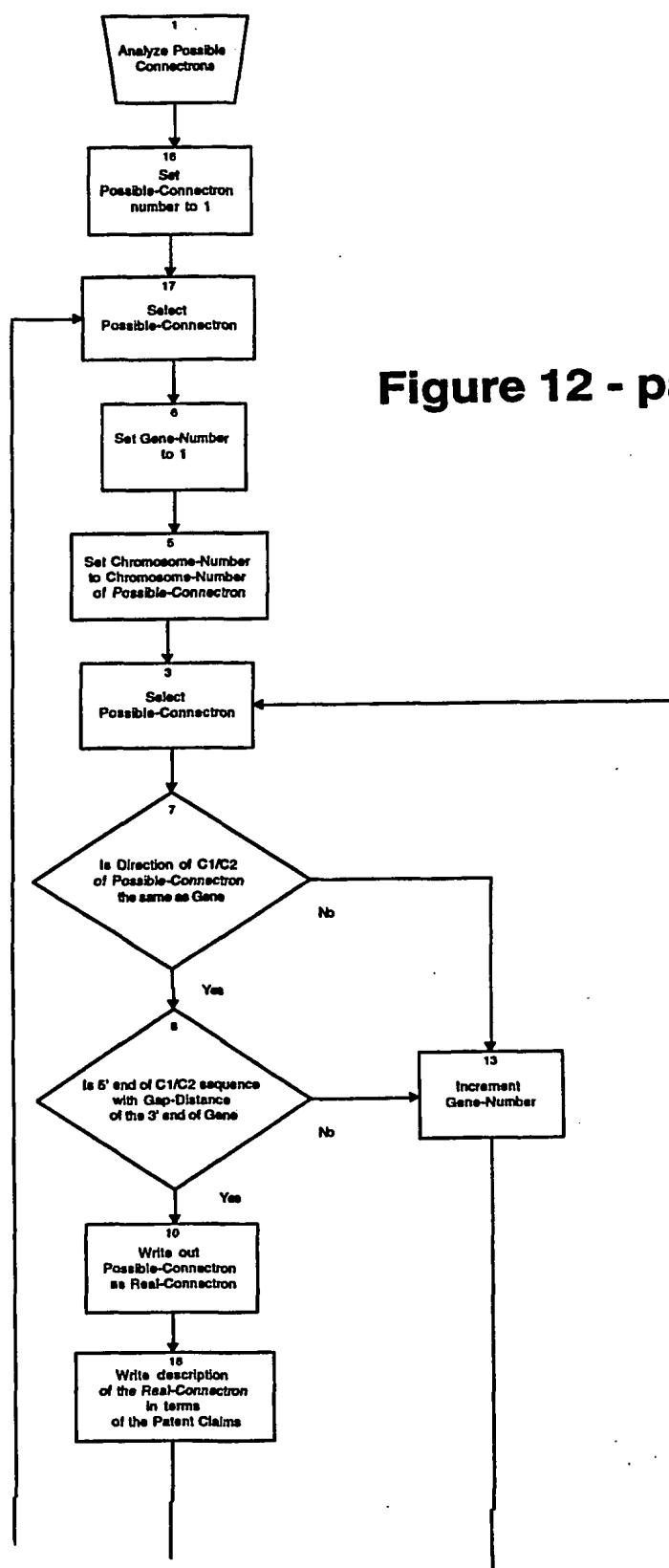


Figure 12 - page 1

19/26

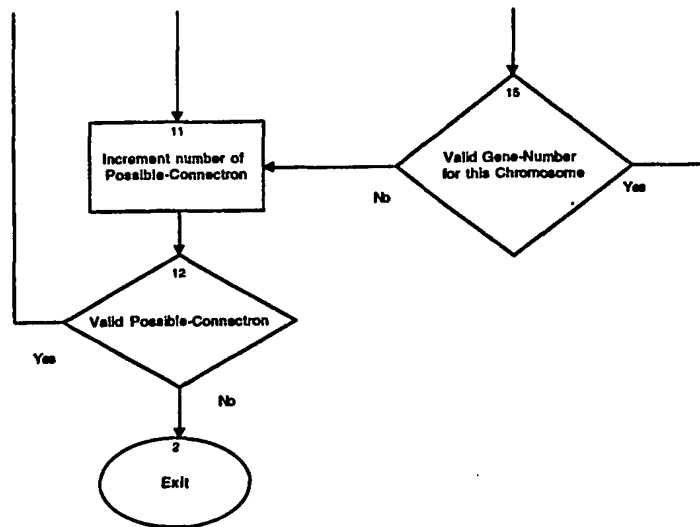


Figure 12 - page 2

20/26

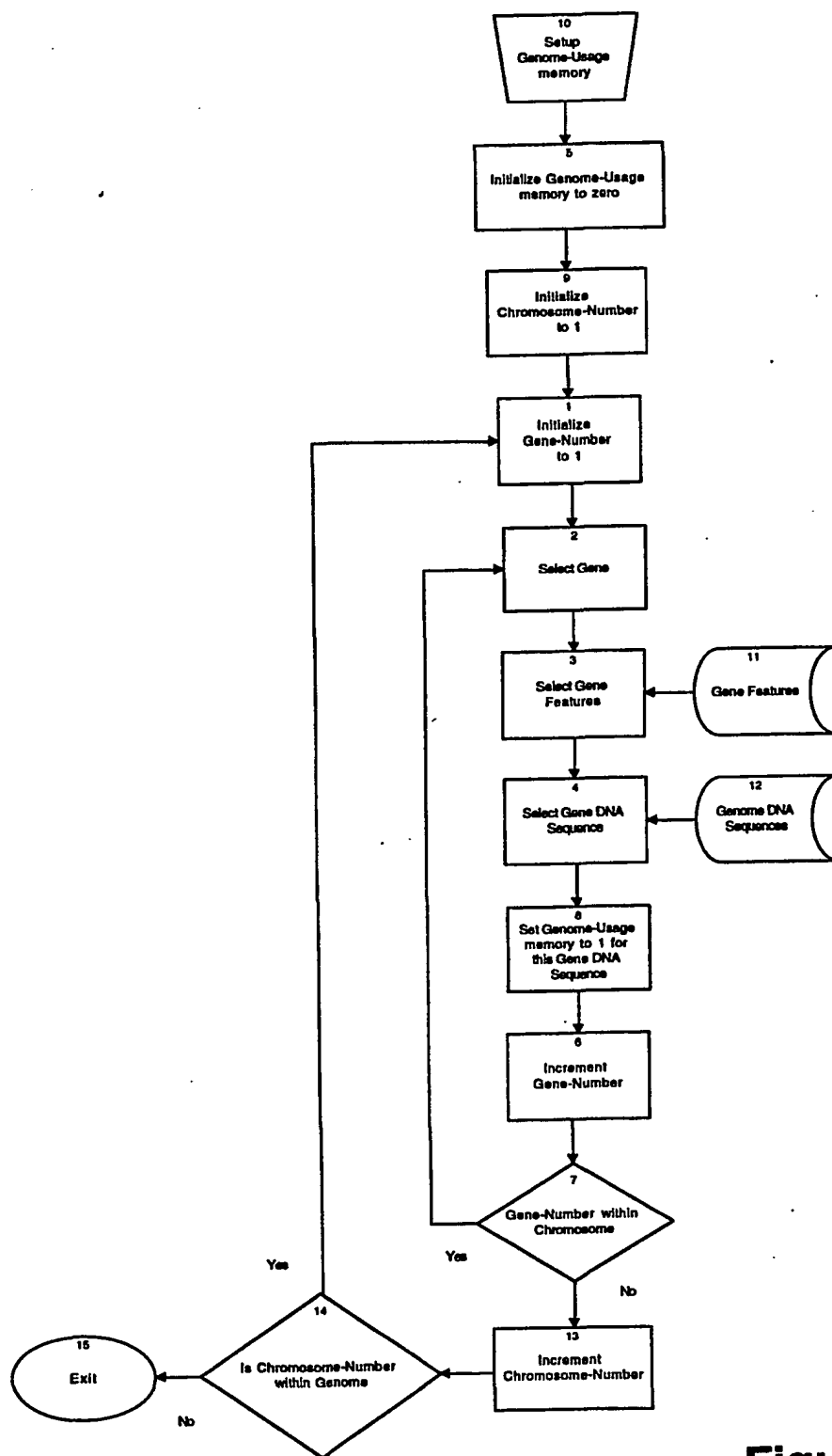


Figure 13

21/26

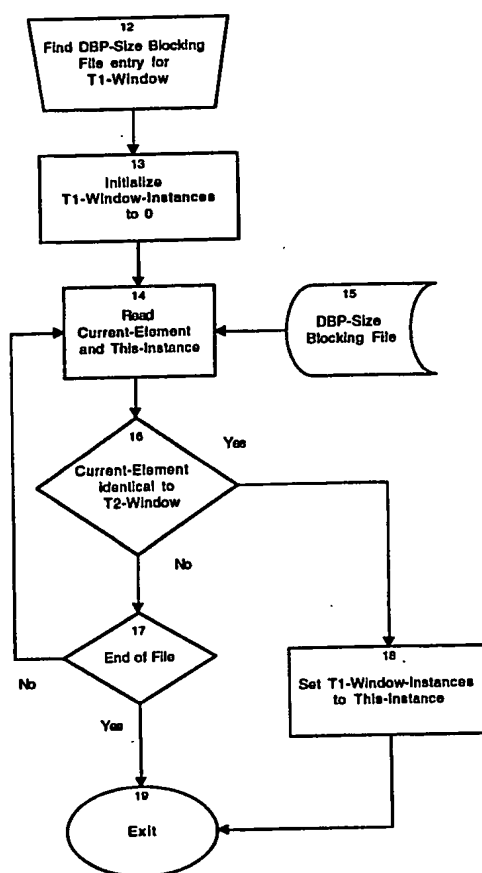


Figure 14

22/26

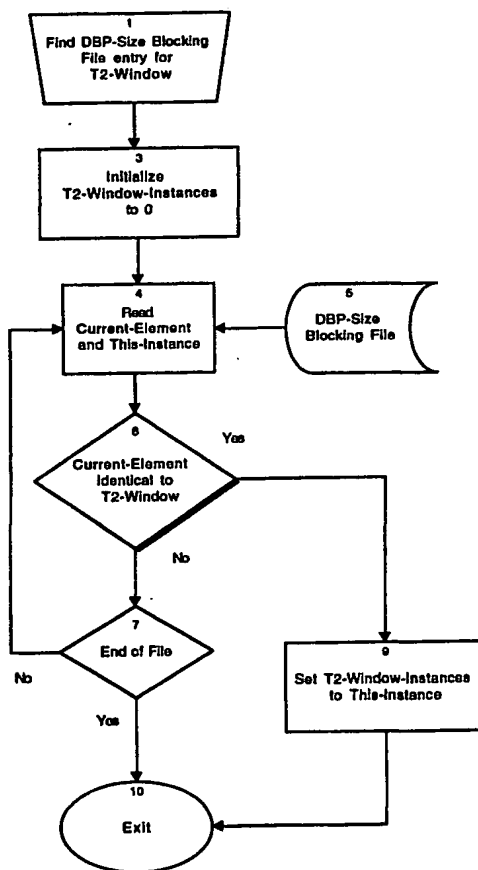
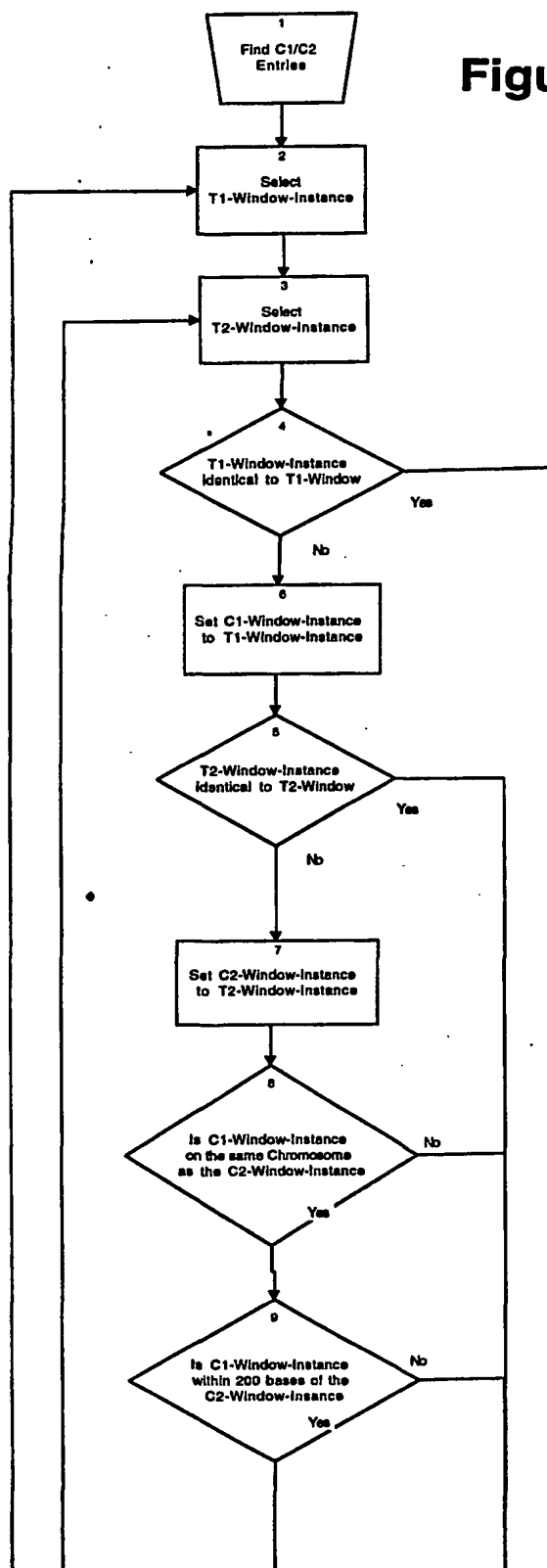


Figure 15

23/26

Figure 16 - page 1



24/26

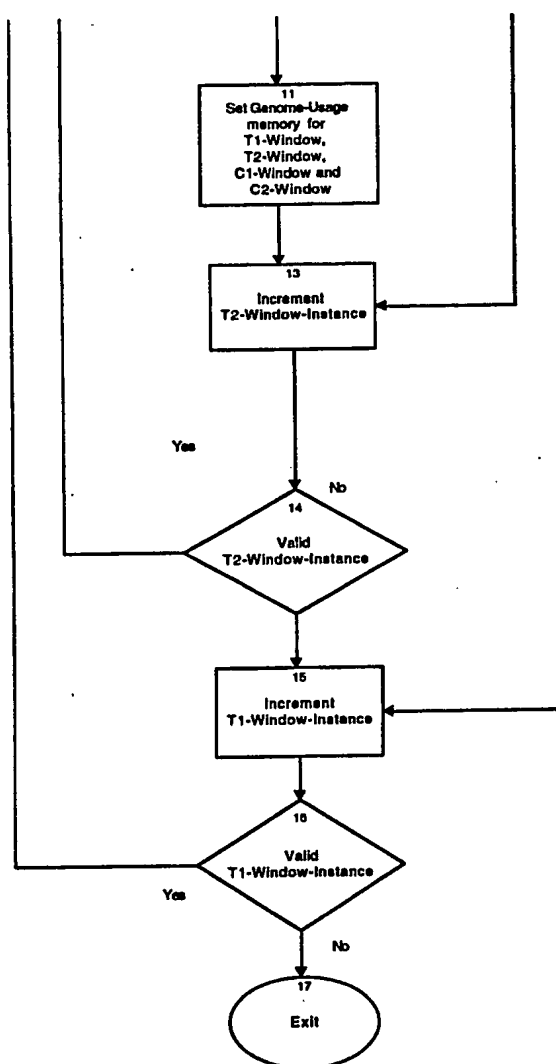
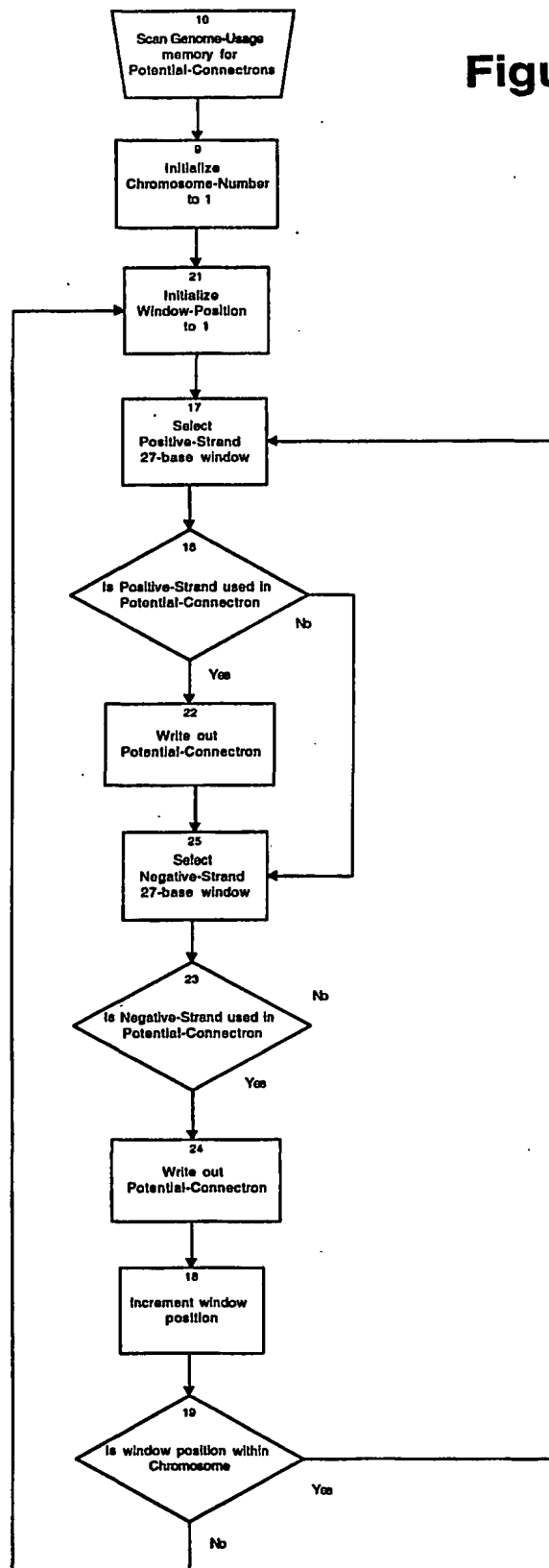


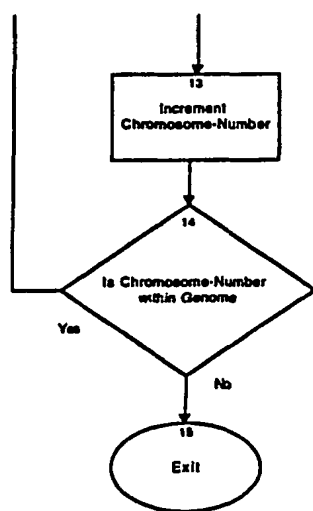
Figure 16 - page 2

25/26

Figure 17 - page 1



26/26

**Figure 17 - page 2**

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
13 December 2001 (13.12.2001)

PCT

(10) International Publication Number
WO 01/94542 A3

- (51) International Patent Classification⁷: **G06F 19/00** (74) Agent: **ZEGEER, Jim**; 801 North Pitt Street #108, Alexandria, VA 22314 (US).
- (21) International Application Number: **PCT/US01/16471**
- (22) International Filing Date: **31 May 2001 (31.05.2001)** (81) Designated States (*national*): **AU, CA, IL, JP, MX.**
- (25) Filing Language: **English** (84) Designated States (*regional*): **European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).**
- (26) Publication Language: **English**
- (30) Priority Data:
60/208,650 2 June 2000 (02.06.2000) US
09/866,925 30 May 2001 (30.05.2001) US
- (71) Applicant: **GLOBAL DETERMINANTS, INC.** [US/US]; 17800 Mill Creek Drive, Derwood, MD 20855-1019 (US).
- (88) Date of publication of the international search report: **18 April 2002**
- (72) Inventor: **FELDMANN, Richard, J.**; 17800 Mill Creek Drive, Derwood, MD 20855-1019 (US).
- Published:
— with international search report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 01/94542 A3

(54) Title: **ALGORITHMIC DETERMINATION OF CONNECTRONS**

(57) Abstract: An algorithm has been developed to identify four DNA sequences of 20 bases or more that form a structure called a connectron. Two sequences C1 and C2 are expressed as RNA in the 3'UTR of some genes in many prokaryotic, archaea and eukaryotic genomes. The other half of a connectron is two DNA sequences T1 and T2 that are 1kb to 105kb apart on the same chromosome. The C1 sequence is identical to the T1 sequence and the C2 sequence is identical to the T2 sequence. C1/C2 and T1-T2 can be on different chromosomes. The C1/C2 RNA sequence of the gene transcript forms a triple-stranded Hoogsteen helix with the double-stranded T1 and T2 DNA sequences. The formation of connectrons blocks expression of genes between T1 and T2.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/16471

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06F 19/00
US CL : 702/19

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 702/19

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KAIN et al. in Current Protocols in Molecular Biology, Ausubel et al. Eds. New York: John Wiley and Sons. 1996, pages 9.6.1 to 9.6.12, especially page 9.6.3.	1
X	ZHANG et al. Identification of Human Gene core Promoters in Silico. Genome Research. 1998, Vol. 8, pages 319-326, especially page 319.	2
A	Matzke et al. How and Why Do Plants Inactivate Homologous (Trans)genes? Plant Physiology. 1995, Vol. 107, pages 679-685, especially page 681.	3-18
A	Fire et al. Potent and specific genetic interference by double-stranded RNA in Caenorabditis elegans. Nature. 19 February 1998, Vol. 391, pages 806-811, especially pages 806-807.	3-18

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	"&" document member of the same patent family

Date of the actual completion of the international search

15 August 2001 (15.08.2001)

Date of mailing of the international search report

23 JAN 2002

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

John S. Brusca, Ph.D.

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/16471

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☒
☐

- The additional search fees were accompanied by the applicant's protest.
No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/16471

Continuation of Item 4 of the first sheet:

The title does not comply with PCT Rule 4.3 because it is longer than 7 words. The new title is as follows:

Algorithmic determination of connectrons.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

- Group I, claim(s) 1, drawn to a promoter detection comprising exclusion of introns and exons.
- Group II, claim(s) 2, drawn to a promoter detection method comprising detection of DNA linked to RNA.
- Group III, claim(s) 3, drawn to a promoter detection method comprising detection of connectron behavior.
- Group IV, claim(s) 4, drawn to a method of modifying expression comprising detection of changes in connectron behavior.
- Group V, claim(s) 5, drawn to a method of detecting integration by detection of connectrons.
- Group VI, claim(s) 6, drawn to a method of detecting expression of genes by detecting flow of connectrons.
- Group VII, claim(s) 7, drawn to a method of modifying a body by modification of connectrons.
- Group VIII, claim(s) 8, drawn to a method of detection of connectrons by any of three methods: a) determining base composition, b) determining one of more sites of control sequence organization, and c) determining sites of target application.
- Group IX, claim(s) 9, drawn to a method of determining cell response by use of complete genome sequences and detection of changes in connectrons caused by stimulus to the genome.
- Group X, claim(s) 10, drawn to polynucleotides with a defined symmetry.
- Group XI, claim(s) 11, drawn to polynucleotides with a first connectron relationship.
- Group XII, claim(s) 12, drawn to polynucleotides with a second connectron relationship.
- Group XIII, claim(s) 13, drawn to polynucleotides with a third connectron relationship.
- Group XIV, claim(s) 14, drawn to polynucleotides with a fourth connectron relationship of plants and higher animals.
- Group XV, claim(s) 15, drawn to polynucleotides with a fifth connectron relationship.
- Group XVI, claim(s) 16, drawn to polynucleotides with a sixth connectron relationship.
- Group XVII, claim(s) 17, drawn to polynucleotides with a seventh connectron relationship.
- Group XVIII, claim(s) 18, drawn to polynucleotides with an eighth connectron relationship.
- Group XIX, claim(s) 19, drawn to polynucleotides with a ninth connectron relationship.
- This application contains claims directed to more than one species of the generic invention. These species are deemed to lack unity of invention because they are not so linked as to form a single general inventive concept under PCT Rule 13.1.
- In order for more than one species to be examined, the appropriate additional examination fees must be paid. The species are as follows:
- Group VIII is drawn to three species of detection of connectrons comprising the alternative steps of: 1) determining base composition, 2) determining one or more sites of control sequence organization, and 3) determining sites of target application.

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/16471

The claims are deemed to correspond to the species listed above in the following manner:

The following claim(s) are generic: 8.

The inventions listed as Groups 1-19 do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: PCT Rule 13.1 and Annex B do not provide for unity of invention between two or more different products or methods of use that share a special technical feature. To the extent the groups have a special technical feature of a connectron, the groups are drawn to different methods of use or detection of connectrons. To the extent the groups are drawn to compositions of connectrons, the groups are drawn to compositions with different structures that lack a common special technical feature.

The species listed above do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: The species listed for Group VIII are drawn to three different methods with different steps that produce different results.

Continuation of B. FIELDS SEARCHED Item 3:
Medline, Biosis, US Patent, Derwent World Patent Index
search terms: promoter, software, connectron